

Automated Feature Engineering

Feature Engineering

- Requires manually writing code

Blocks

- Time Consuming Process

Feature Engineering

- Requires manually writing code

Blocks

- Time Consuming Process



Automated Feature Engineering

- Requires manually writing code

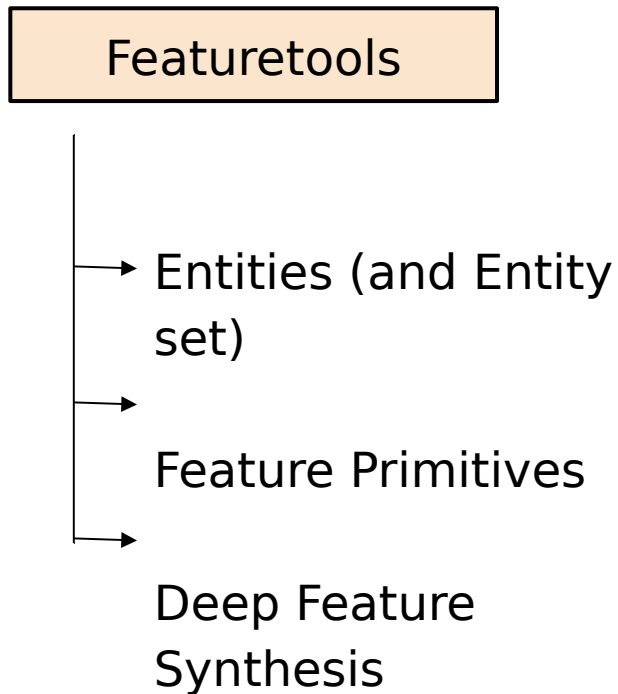
Blocks

Featuretools

- Time Consuming Process



Automated Feature Engineering



Automated Feature Engineering

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier	Establish Year	Outlet Size	Location Type	Outlet Type	Item Outlet Sales
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192

Automated Feature Engineering

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier	Establish Year	Outlet Size	Location Type	Outlet Type	Item Outlet Sales
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192

Automated Feature Engineering

● Entity

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier	Establish Year	Outlet Size	Location Type	Outlet Type	Item Outlet Sales
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192

Automated Feature Engineering

● Entity

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier	Establish Year	Outlet Size	Location Type	Outlet Type	Item Outlet Sales
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192

Automated Feature Engineering

● Entity

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier	Establish Year	Outlet Size	Location Type	Outlet Type	Item Outlet Sales
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.138
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.27
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010	1998		Tier 3	Grocery Store	732.38
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018	2009	Medium	Tier 3	Supermarket Type2	556.6088
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013	1987	High	Tier 3	Supermarket Type1	343.5528
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027	1985	Medium	Tier 3	Supermarket Type3	4022.7636
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045	2002		Tier 2	Supermarket Type1	1076.5986
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017	2007		Tier 2	Supermarket Type1	4710.535
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049	1999	Medium	Tier 1	Supermarket Type1	1516.0266
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046	1997	Small	Tier 1	Supermarket Type1	2187.153
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049	1999	Medium	Tier 1	Supermarket Type1	1589.2646
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046	1997	Small	Tier 1	Supermarket Type1	2145.2076
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013	1987	High	Tier 3	Supermarket Type1	1977.426
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046	1997	Small	Tier 1	Supermarket Type1	1547.3192

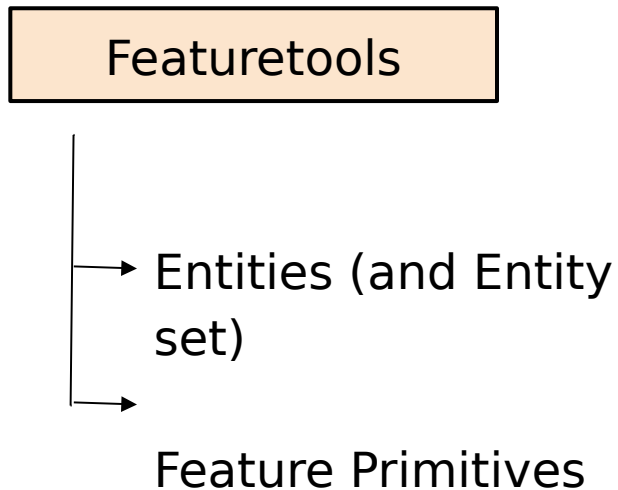
Automated Feature Engineering

- Entity : A table or a dataframe

Automated Feature Engineering

- Entity : A table or a dataframe
- Entity Set: A collection of tables or entities

Automated Feature Engineering



Automated Feature Engineering

Featuretools

→ Entities (and Entity set)

→ Feature Primitives

name	type	description
num_true	aggregation	Finds the number of 'True' values in a boolean.
percent_true	aggregation	Finds the percent of 'True' values in a boolean feature.
time_since_last	aggregation	Time since last related instance.
num_unique	aggregation	Returns the number of unique categorical variables.
avg_time_between	aggregation	Computes the average time between consecutive events.
all	aggregation	Test if all values are 'True'.
min	aggregation	Finds the minimum non-null value of a numeric feature.
mean	aggregation	Computes the average value of a numeric feature.
seconds	transform	Transform a Timedelta feature into the number of seconds.
second	transform	Transform a Datetime feature into the second.
and	transform	For two boolean values, determine if both values are 'True'.
month	transform	Transform a Datetime feature into the month.
cum_sum	transform	Calculates the sum of previous values of an instance for each value in a time-dependent entity.
percentile	transform	For each value of the base feature, determines the percentile in relation
time_since_previous	transform	Compute the time since the previous instance.
cum_min	transform	Calculates the min of previous values of an instance for each value in a time-dependent entity.

Automated Feature Engineering

Item_weight + Item_MRP

Item_weight * Item_MRP

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046

Automated Feature Engineering

Item_weight + Item_MRP

Item_weight * Item_MRP

Item_weight * Item_MRP +
Item_weight

Item Identifier	Item Weight	Item Fat Content	Item Visibility	Item Type	Item MRP	Outlet Identifier
FDA15	9.3	Low Fat	0.01604730	Dairy	249.8092	OUT049
DRC01	5.92	Regular	0.01927821	Soft Drinks	48.2692	OUT018
FDN15	17.5	Low Fat	0.01676007	Meat	141.618	OUT049
FDX07	19.2	Regular	0	Fruits and Vegetables	182.095	OUT010
NCD19	8.93	Low Fat	0	Household	53.8614	OUT013
FDP36	10.395	Regular	0	Baking Goods	51.4008	OUT018
FDO10	13.65	Regular	0.01274108	Snack Foods	57.6588	OUT013
FDP10		Low Fat	0.12746985	Snack Foods	107.7622	OUT027
FDH17	16.2	Regular	0.01668711	Frozen Foods	96.9726	OUT045
FDU28	19.2	Regular	0.09444959	Frozen Foods	187.8214	OUT017
FDY07	11.8	Low Fat	0	Fruits and Vegetables	45.5402	OUT049
FDA03	18.5	Regular	0.04546377	Dairy	144.1102	OUT046
FDX32	15.1	Regular	0.1000135	Fruits and Vegetables	145.4786	OUT049
FDS46	17.6	Regular	0.04725732	Snack Foods	119.6782	OUT046
FDF32	16.35	Low Fat	0.0680243	Fruits and Vegetables	196.4426	OUT013
FDP49	9	Regular	0.06908896	Breakfast	56.3614	OUT046

Automated Feature Engineering

