NAME:S.S.NITHISH SOORYA

REG NO:19BEC0212

CAMPUS : VELLORE

DATA ANALYTICS

ASSIGNMENT - 3

CLASSIFICATION OF TERM DEPOSIT BUYERS FOR RETAIL BANKING

IBM Watson Studio | Search in your workspaces | Buy | NITHISH SOORYA S S's Ac... | NS

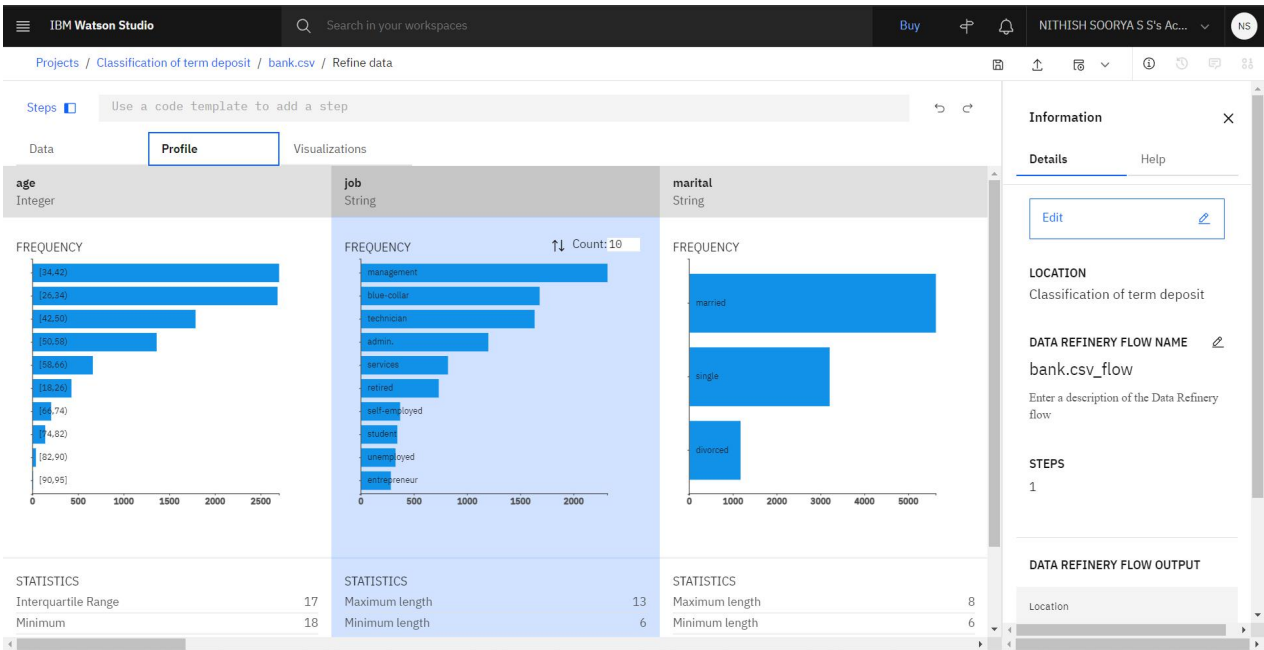My projects / Classification of term deposit / bank.csv
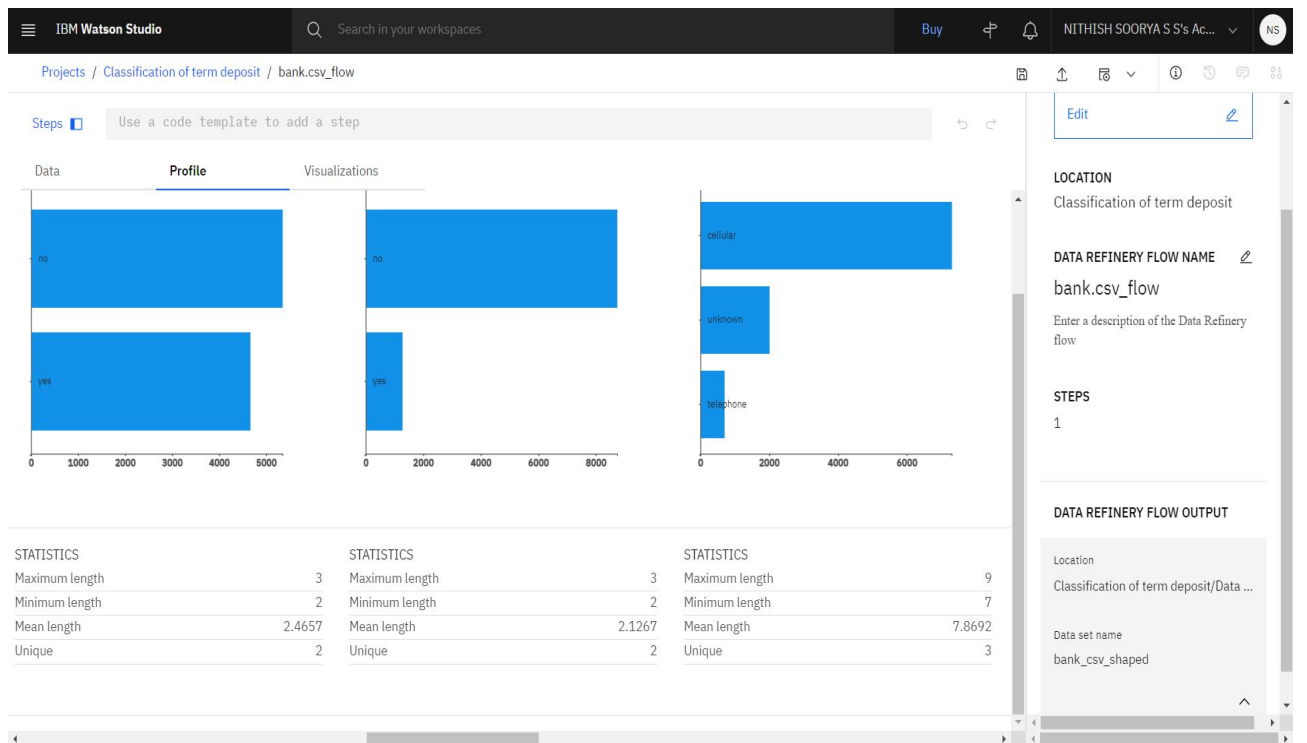
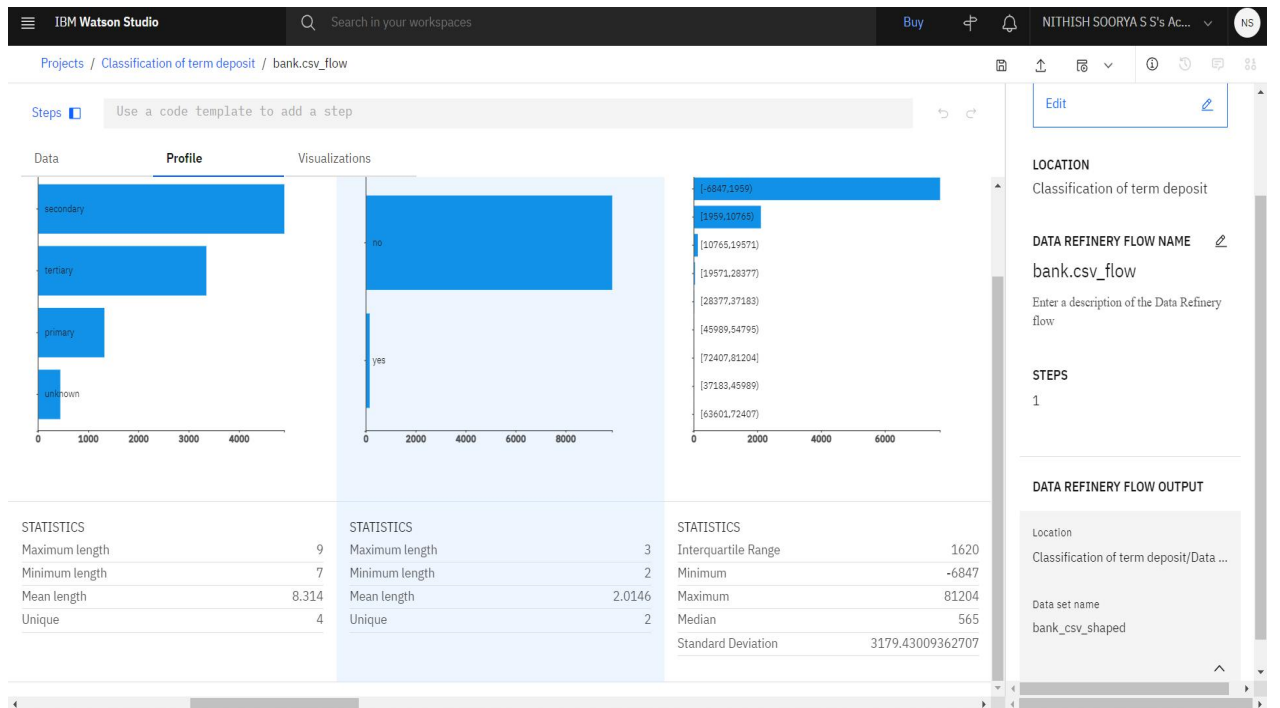Preview | Activities

Schema: 17 Columns
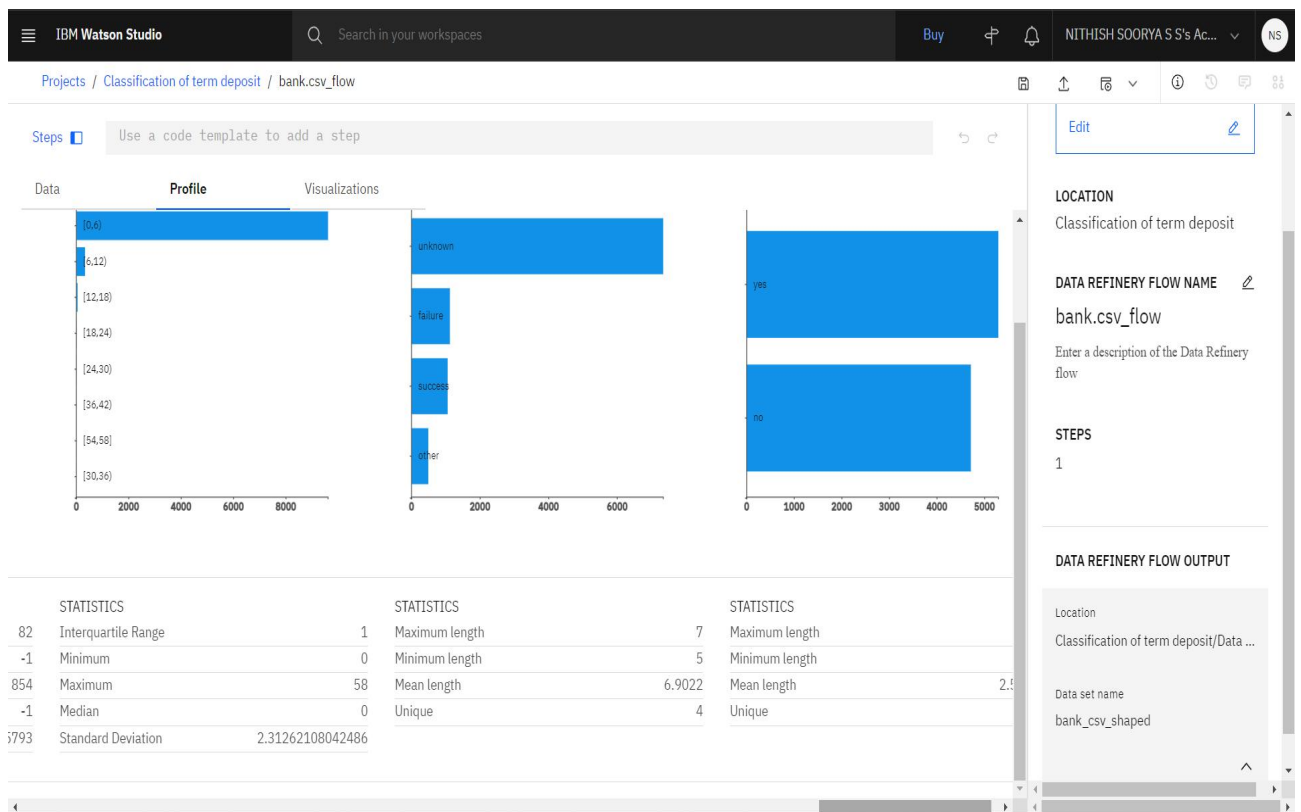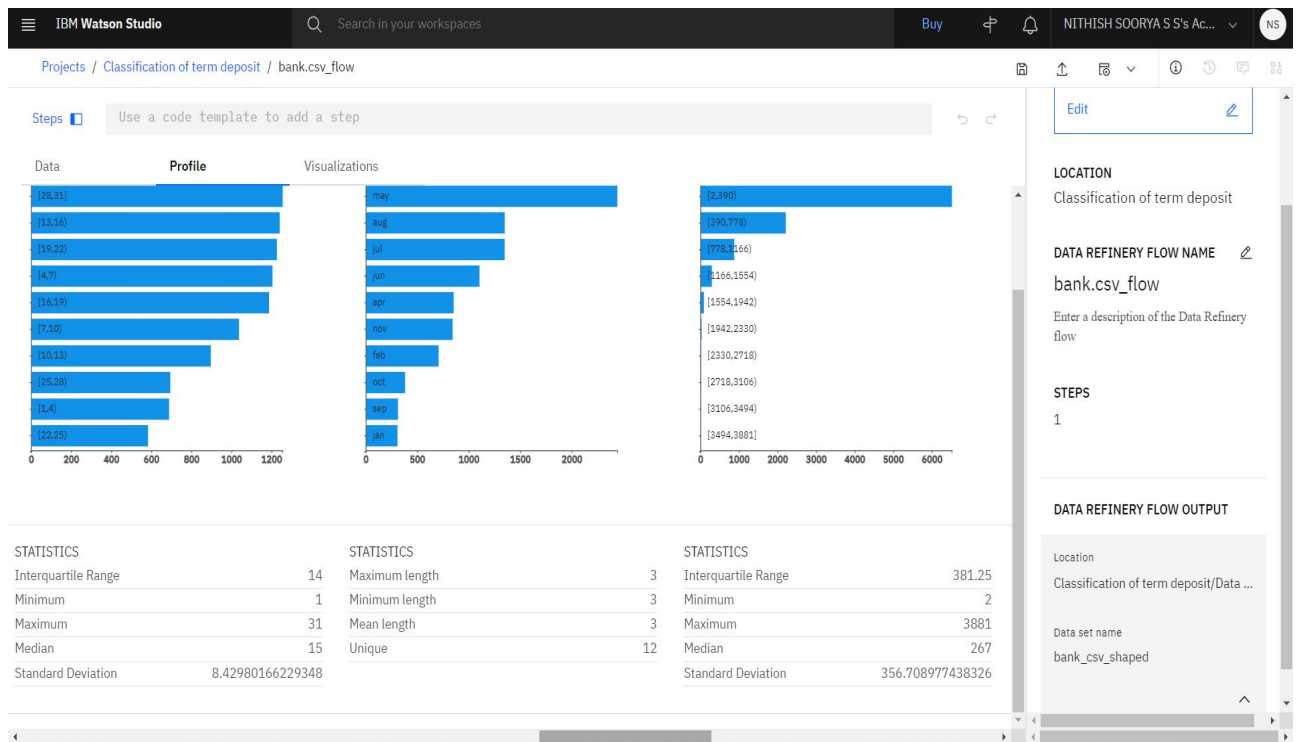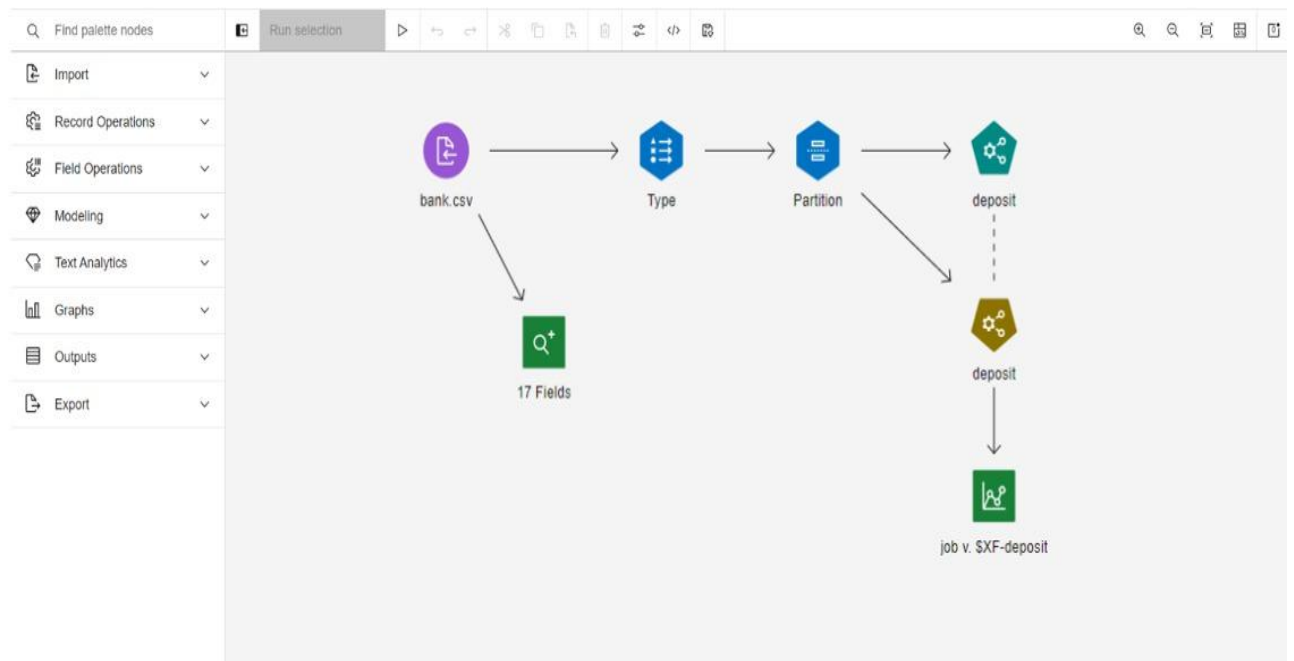The preview includes only a limited set of columns and rows.

Last refresh: just now | Refine

age String	job String	marital String	education String	default String	balance String	housing String	loan String	contact String	day String	month String	duration String	campaign String
59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1
56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1
41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1
55	services	married	secondary	no	2476	yes	no	unknown	5	may	579	1
54	admin.	married	tertiary	no	184	no	no	unknown	5	may	673	2
42	management	single	tertiary	no	0	yes	yes	unknown	5	may	562	2
56	management	married	tertiary	no	830	yes	yes	unknown	6	may	1201	1
60	retired	divorced	secondary	no	545	yes	no	unknown	6	may	1030	1
37	technician	married	secondary	no	1	yes	no	unknown	6	may	608	1
28	services	single	secondary	no	5090	yes	no	unknown	6	may	1297	3

IBM Watson Studio | Search in your workspaces | Buy | NITHISH SOORYA S S's Ac... | NS

Projects / Classification of term deposit / bank.csv / Refine data

Steps | Use a code template to add a step

Data | Profile | Visualizations

Information | X

Details | Help

age Integer | **job** String | **marital** String

FREQUENCY | FREQUENCY Count: 10 | FREQUENCY

Edit

LOCATION
Classification of term deposit

DATA REFINERY FLOW NAME
bank.csv_flow
Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Location

STATISTICS
Interquartile Range	17	Maximum length	13	Maximum length	8
Minimum	18	Minimum length	6	Minimum length	6

Projects / Classification of term deposit / bank.csv_flow

Steps

Use a code template to add a step

Data   Profile   Visualizations

secondary

tertiary

primary

unknown

0   1000   2000   3000   4000

no

yes

0   2000   4000   6000   8000

[-6847,1959)
[1959,10765)
[10765,19571)
[19571,28377)
[28377,37183)
[45989,54795)
[72407,81204)
[37183,45989)
[63601,72407)

0   2000   4000   6000

Edit

LOCATION
Classification of term deposit

DATA REFINERY FLOW NAME
bank.csv_flow

Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Location
Classification of term deposit/Data ...

Data set name
bank_csv_shaped

STATISTICS		STATISTICS		STATISTICS	
Maximum length	9	Maximum length	3	Interquartile Range	1620
Minimum length	7	Minimum length	2	Minimum	-6847
Mean length	8.314	Mean length	2.0146	Maximum	81204
Unique	4	Unique	2	Median	565
				Standard Deviation	3179.43009362707

Projects / Classification of term deposit / bank.csv_flow

Steps

Use a code template to add a step

Data   Profile   Visualizations

no

yes

0   1000   2000   3000   4000   5000

no

yes

0   2000   4000   6000   8000

cellular

unknown

telephone

0   2000   4000   6000

Edit

LOCATION
Classification of term deposit

DATA REFINERY FLOW NAME
bank.csv_flow

Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Location
Classification of term deposit/Data ...

Data set name
bank_csv_shaped

STATISTICS		STATISTICS		STATISTICS	
Maximum length	3	Maximum length	3	Maximum length	9
Minimum length	2	Minimum length	2	Minimum length	7
Mean length	2.4657	Mean length	2.1267	Mean length	7.8692
Unique	2	Unique	2	Unique	3

Steps   Use a code template to add a step

Data   **Profile**   Visualizations

| [28,31] |
| [13,16] |
| [19,22] |
| [4,7] |
| [16,19] |
| [7,10] |
| [10,13] |
| [25,28] |
| [1,4] |
| [22,25] |

0 200 400 600 800 1000 1200

| may |
| aug |
| jul |
| jun |
| apr |
| nov |
| feb |
| oct |
| sep |
| jan |

0 500 1000 1500 2000

| [2,390) |
| [390,778) |
| [778,1166) |
| [1166,1554) |
| [1554,1942) |
| [1942,2330) |
| [2330,2718) |
| [2718,3106) |
| [3106,3494) |
| [3494,3881] |

0 1000 2000 3000 4000 5000 6000

STATISTICS

Interquartile Range	14
Minimum	1
Maximum	31
Median	15
Standard Deviation	8.42980166229348

STATISTICS

Maximum length	3
Minimum length	3
Mean length	3
Unique	12

STATISTICS

Interquartile Range	381.25
Minimum	2
Maximum	3881
Median	267
Standard Deviation	356.708977438326

Edit

LOCATION
Classification of term deposit

DATA REFINERY FLOW NAME
bank.csv_flow
Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Location
Classification of term deposit/Data ...

Data set name
bank_csv_shaped

Steps   Use a code template to add a step

Data   **Profile**   Visualizations

| [0,6) |
| [6,12) |
| [12,18) |
| [18,24) |
| [24,30) |
| [36,42) |
| [54,58) |
| [30,36) |

0 2000 4000 6000 8000

| unknown |
| failure |
| success |
| other |

0 2000 4000 6000

| yes |
| no |

0 1000 2000 3000 4000 5000

STATISTICS

82	Interquartile Range	1
-1	Minimum	0
854	Maximum	58
-1	Median	0
5793	Standard Deviation	2.31262108042486

STATISTICS

Maximum length	7
Minimum length	5
Mean length	6.9022
Unique	4

STATISTICS

Maximum length	
Minimum length	
Mean length	2.5
Unique	

Edit

LOCATION
Classification of term deposit

DATA REFINERY FLOW NAME
bank.csv_flow
Enter a description of the Data Refinery flow

STEPS
1

DATA REFINERY FLOW OUTPUT

Location
Classification of term deposit/Data ...

Data set name
bank_csv_shaped

Projects / Classification of term deposit / classification of term deposit

View Output: Data Audit of [17 fields]

Compare

	Field	Sample Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness	Unique	Valid
1	age		Continuous	18	95	41.232	11.913	0.863	--	11162
2	job		Categorical	--	--	--	--	--	12	11162
3	marital		Categorical	--	--	--	--	--	3	11162
4	education		Categorical	--	--	--	--	--	4	11162
5	default		Categorical	--	--	--	--	--	2	11162
6	balance		Continuous	-6847	81204	1528.539	3225.413	8.225	--	11162
7	housing		Categorical	--	--	--	--	--	2	11162

View Output: Data Audit of [17 fields]

Compare ⊞   ✕

	Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String
1	age	Continuous	132	0	None	Never	Fixed	100.000	11162	0	0
2	job	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
3	marital	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
4	education	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
5	default	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
6	balance	Continuous	99	74	None	Never	Fixed	100.000	11162	0	0
7	housing	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
8	loan	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
9	contact	Categorical	--	--	--	Never	Fixed	100.000	11162	0	0
10	day	Continuous	0	0	None	Never	Fixed	100.000	11162	0	0

Q Find in column Field

	Field	Measure		Role		Value mode	Values	Check	
☐	# age	Continuous	∨	Input	∨	Instantiated ∨	18, 95	None	∨ ⚙
☐	abc job	Nominal	∨	Input	∨	Instantiated ∨	admin., blue-collar,...	None	∨ ⚙
☐	abc marital	Nominal	∨	Input	∨	Instantiated ∨	divorced, married, ...	None	∨ ⚙
☐	abc education	Nominal	∨	Input	∨	Instantiated ∨	primary, secondary...	None	∨ ⚙
☐	abc default	Flag	∨	None	∨	Instantiated ∨	no, yes	None	∨ ⚙
☐	# balance	Continuous	∨	Input	∨	Instantiated ∨	-6847, 81204	None	∨ ⚙
☐	abc housing	Flag	∨	Input	∨	Instantiated ∨	no, yes	None	∨ ⚙
☐	abc loan	Flag	∨	Input	∨	Instantiated ∨	no, yes	None	∨ ⚙

	Field	Measure		Role		Value mode	Values	Check	
☐	# day	Continuous	⌄	Input	⌄	Instantiated ⌄	1, 31	None ⌄	⚙
☐	abc month	Nominal	⌄	Input	⌄	Instantiated ⌄	apr, aug, dec, feb, …	None ⌄	⚙
☐	# duration	Continuous	⌄	Input	⌄	Instantiated ⌄	2, 3881	None ⌄	⚙
☐	# campaign	Continuous	⌄	Input	⌄	Instantiated ⌄	1, 63	None ⌄	⚙
☐	# pdays	Continuous	⌄	Input	⌄	Instantiated ⌄	-1, 854	None ⌄	⚙
☐	# previous	Continuous	⌄	Input	⌄	Instantiated ⌄	0, 58	None ⌄	⚙
☐	abc poutcome	Nominal	⌄	None	⌄	Instantiated ⌄	failure, other, succ…	None ⌄	⚙
☐	abc deposit	Flag	⌄	Target	⌄	Instantiated ⌄	no, yes	None ⌄	⚙

Q Find in column Field

View Output: Analysis of [deposit]

⊖ **Results for output field deposit**

⊖ **Comparing $XF-deposit with deposit**

'Partition'	1_Training		2_Testing	
Correct	6,918	88.81%	2,857	84.73%
Wrong	872	11.19%	515	15.27%
Total	7,790		3,372	

View Model: deposit

Auto Classifier ⓘ

Models

Auto Classifier - Models ⓘ

TARGET : DEPOSIT

USE	MODEL NAME	ESTIMATOR ▼	BUILD TIME (MINS)	NO. FIELDS USED	ACCURACY	ACCUMULATED ACCURACY	AREA UNDER CURVE	ACCUMULATED AUC	RECALL	PRECISION	ACTI
☑	Random Trees 1	Random Trees	1	14	83.096	83.096	0.893	0.893	0.860	0.800	
☑	Logistic regression 1	Nominal Regression	<1	14	80.042	80.042	0.885	0.885	0.765	0.805	
☑	C5 1	C5.0	<1	13	85.024	85.024	0.897	0.897	0.878	0.819	
☑	Bayesian Network 1	BayesNet	<1	14	78.529	78.529	0.835	0.835	0.708	0.819	

0.8973571

View Model: deposit

Auto Classifier ⓘ

Models

Auto Classifier - Models ⓘ

TARGET : DEPOSIT

USE	MODEL NAME	ESTIMATOR ▼	BUILD TIME (MINS)	NO. FIELDS USED	ACCURACY	ACCUMULATED ACCURACY	AREA UNDER CURVE	ACCUMULATED AUC	RECALL	PRECISION
☑	XGBoost Tree 1	XGBoost Binary Classification Model	1	14	83.482	83.482	0.906	0.906	0.846	0.813
☑	Random Trees 1	Random Trees	1	14	83.096	83.096	0.893	0.893	0.860	0.800
☑	Logistic regression 1	Nominal Regression	<1	14	80.042	80.042	0.885	0.885	0.765	0.805
☑	C5 1	C5.0	<1	13	85.024	85.024	0.897	0.897	0.878	0.819

View Output: job v. $XF-deposit