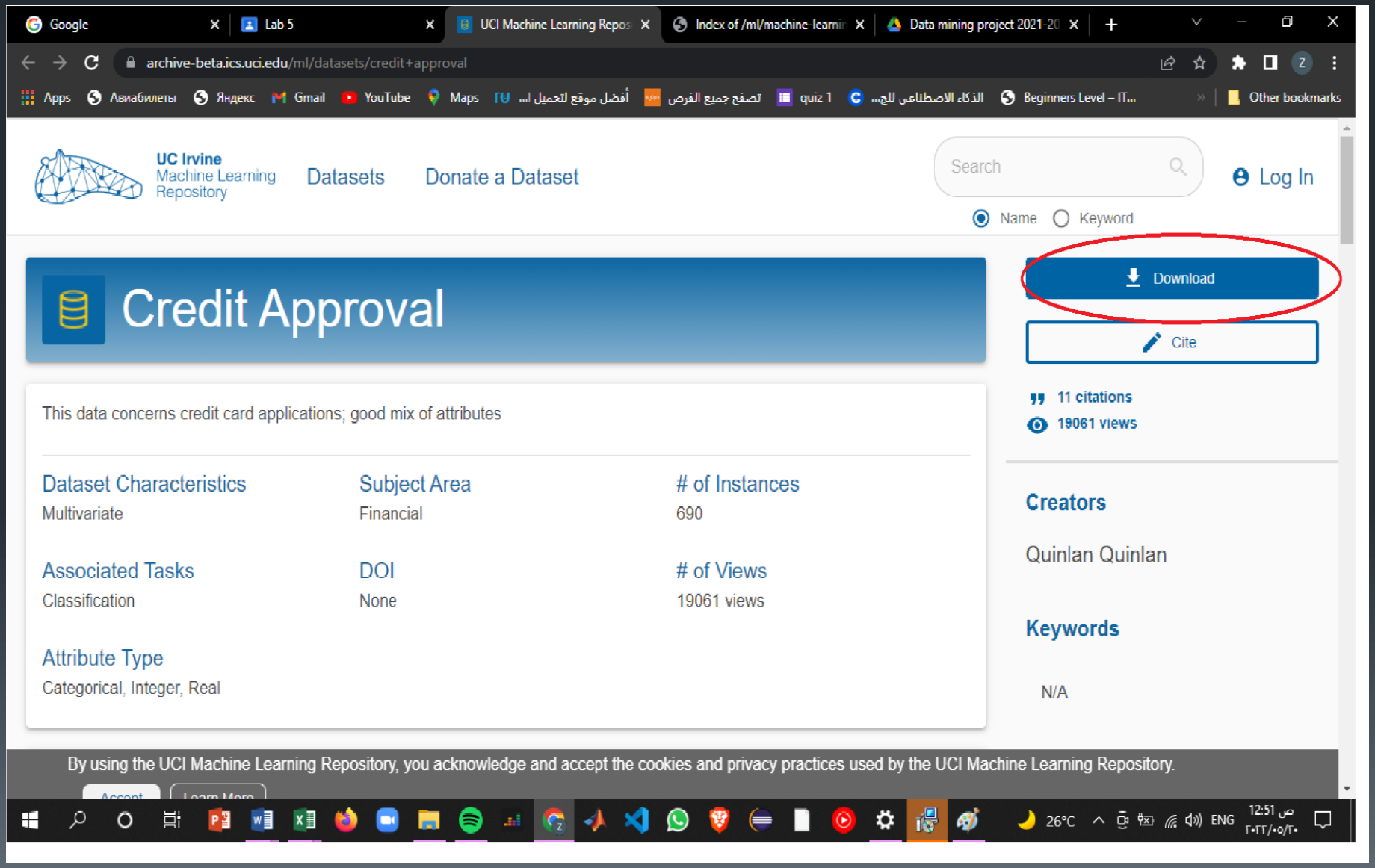
Data Mining Project

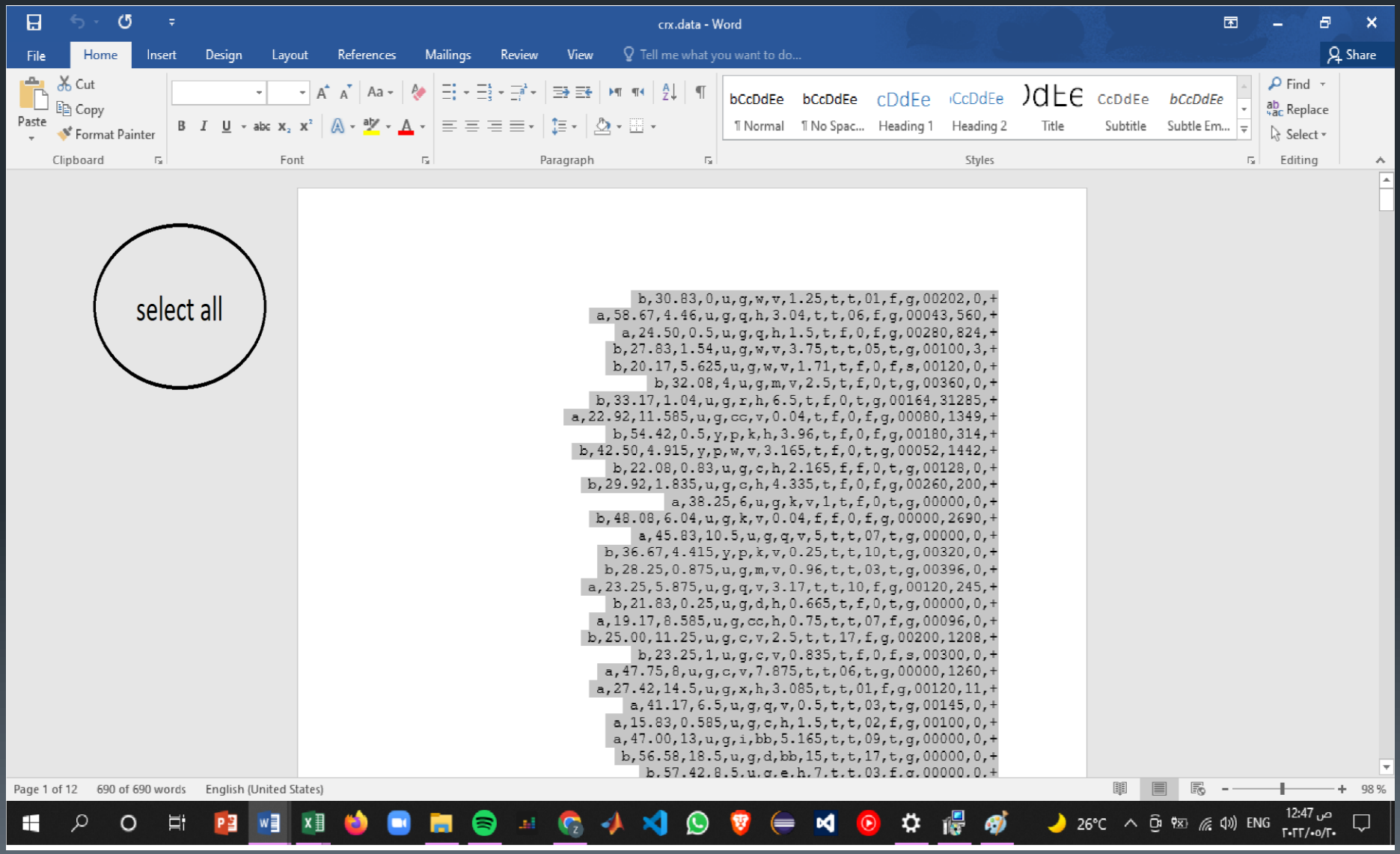
- **Data Set: Credit Approval**
- **Members Names:**

1) Zead Hossam Eldin Samir (CS 2)
2) George Morqus Nagib (CS 2)
3) Islam Atef Nagi (CS 1)
4) Yousef Shaban Mohamed (CS 1)
5) Mohamed Ahmed Mufreh (CS 2)
6) Mohamed Kamal Fathy (CS 2)

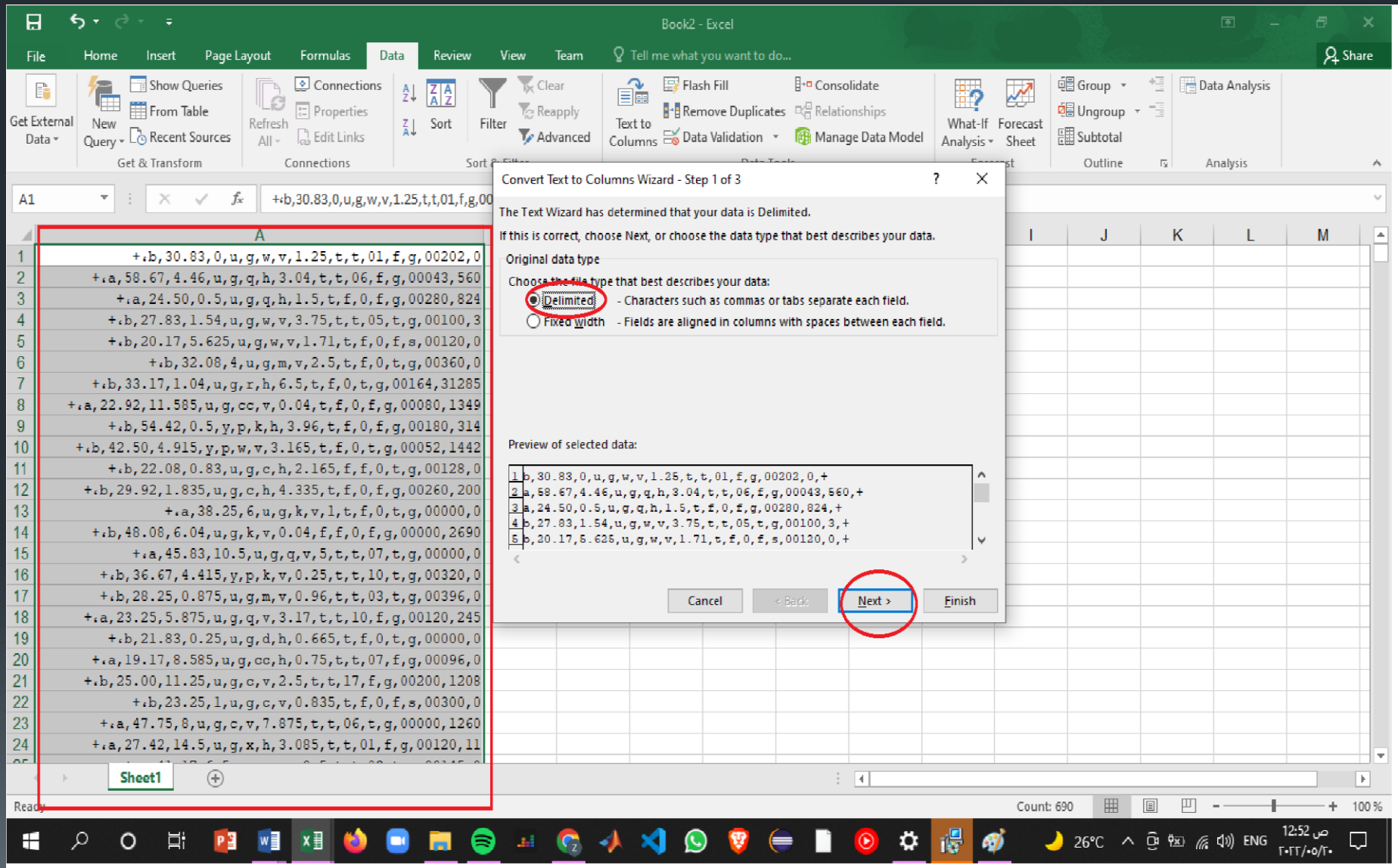
Downloading Dataset



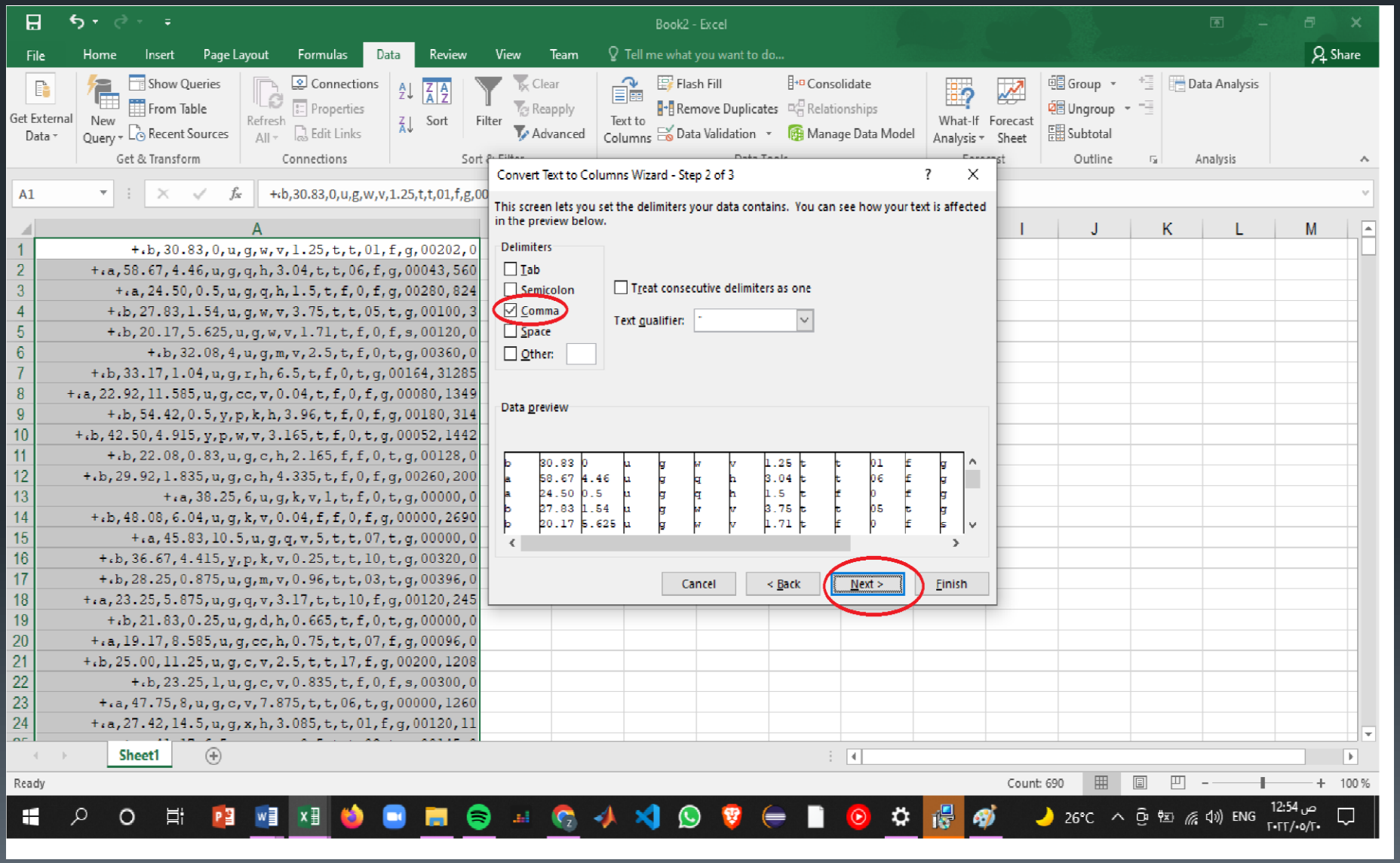
Viewing Dataset



Preparing Dataset



Preparing Dataset cont…



Dataset is Prepared

the data set

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
b	30.83	0	u	g	w	v	1.25	t	t	1	f	g	202	0	+
a	58.67	4.46	u	g	q	h	3.04	t	t	6	f	g	43	560	+
a	24.5	0.5	u	g	q	h	1.5	t	f	0	f	g	280	824	+
b	27.83	1.54	u	g	w	v	3.75	t	t	5	t	g	100	3	+
b	20.17	5.625	u	g	w	v	1.71	t	f	0	f	s	120	0	+
b	32.08	4	u	g	m	v	2.5	t	f	0	t	g	360	0	+
b	33.17	1.04	u	g	r	h	6.5	t	f	0	t	g	164	31285	+
a	22.92	11.585	u	g	cc	v	0.04	t	f	0	f	g	80	1349	+
b	54.42	0.5	y	p	k	h	3.96	t	f	0	f	g	180	314	+
b	42.5	4.915	y	p	w	v	3.165	t	f	0	t	g	52	1442	+
b	22.08	0.83	u	g	c	h	2.165	f	f	0	t	g	128	0	+
b	29.92	1.835	u	g	c	h	4.335	t	f	0	f	g	260	200	+
a	38.25	6	u	g	k	v	1	t	f	0	t	g	0	0	+
b	48.08	6.04	u	g	k	v	0.04	f	f	0	f	g	0	2690	+
a	45.83	10.5	u	g	q	v	5	t	t	7	t	g	0	0	+
b	36.67	4.415	y	p	k	v	0.25	t	t	10	t	g	320	0	+
b	28.25	0.875	u	g	m	v	0.96	t	t	3	t	g	396	0	+
a	23.25	5.875	u	g	q	v	3.17	t	t	10	f	g	120	245	+
b	21.83	0.25	u	g	d	h	0.665	t	f	0	t	g	0	0	+
a	19.17	8.585	u	g	cc	h	0.75	t	t	7	f	g	96	0	+
b	25	11.25	u	g	c	v	2.5	t	t	17	f	g	200	1208	+
b	23.25	1	u	g	c	v	0.835	t	f	0	f	s	300	0	+
a	47.75	8	u	g	c	v	7.875	t	t	6	t	g	0	1260	+

Missing attributes

crx.names - Notepad

File Edit Format View Help

```
    A12:        t, f.
    A13:        g, p, s.
    A14:        continuous.
    A15:        continuous.
    A16: +,-        (class attribute)

8.  Missing Attribute Values:
    37 cases (5%) have one or more missing values.  The missing
    values from particular attributes are:

    A1:  12
    A2:  12
    A4:   6
    A5:   6
    A6:   9
    A7:   9
    A14: 13

9.  Class Distribution

    +: 307 (44.5%)
    -: 383 (55.5%)
```

Ln 1, Col 1 100% Unix (LF) UTF-8

Handling missing values (1)

Handling missing values (2)

Stream1* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Credit Approval.xlsx → Type

Type

Preview

Types Format Annotations

Read Values Clear Values Clear All Values

Field	Measurement	Values	Missing	Check	Role
A1	Nominal	"?",a,b		None	Input
A2	Continuous	[13.75,80....		None	Input
A3	Continuous	[0.0,28.0]		None	Input
A4	Nominal	"?",l,u,y		None	Input
A5	Nominal	"?",g,gg,p		None	Input
A6	Nominal	"?",aa,c,cc...		None	Input
A7	Nominal	"?",bb,dd,f...		None	Input
A8	Continuous	[0.0,28.5]		None	Input
A9	Flag	t/f		None	Input

View current fields View unused field settings

OK Cancel Apply Reset

Streams Outputs Models

Stream1

CRISP-DM Classes

(unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

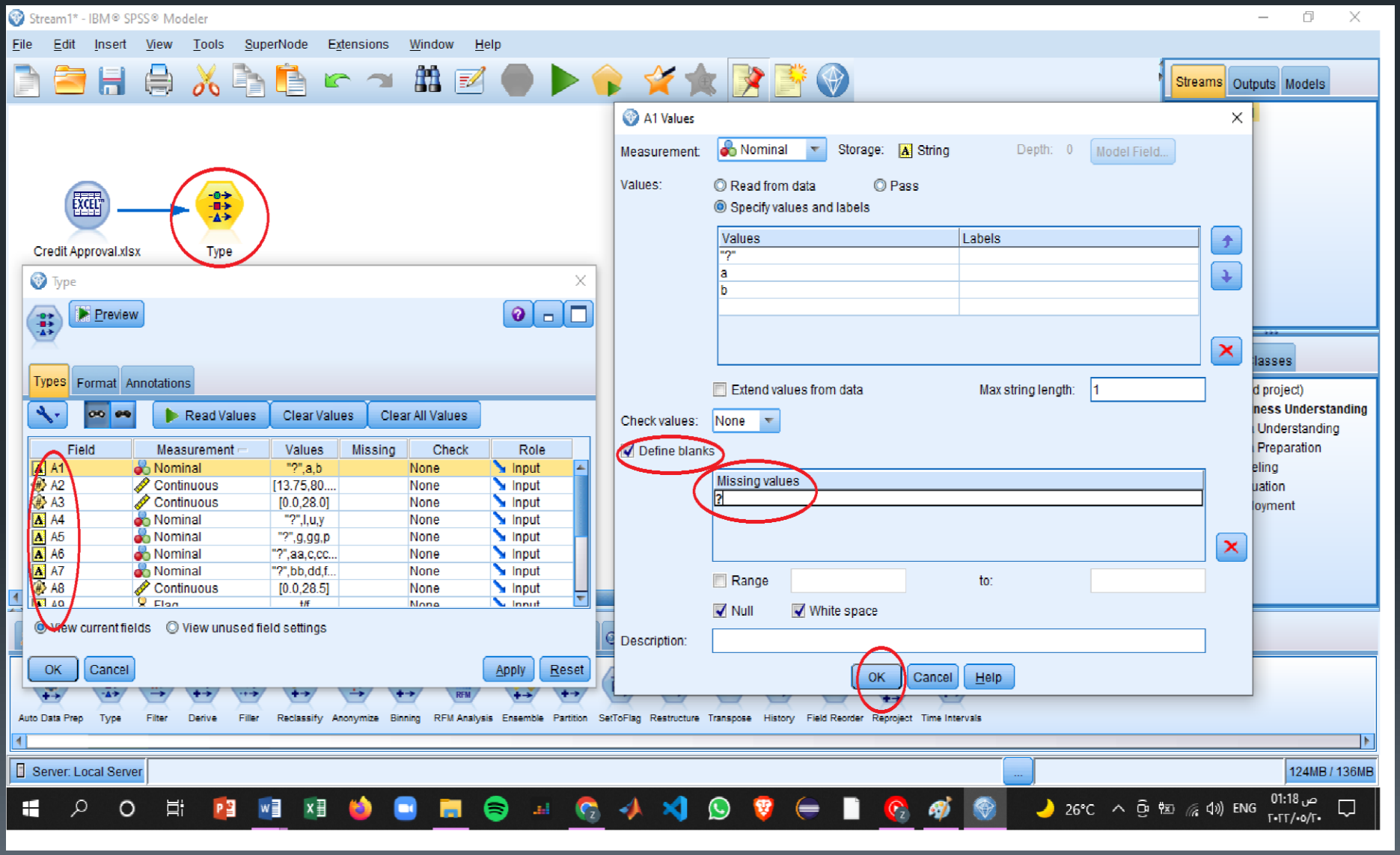
Favorites Sources Record Ops Field Ops Graphs Modeling Output Export IBM® SPSS® Statistics

Auto Data Prep Type Filter Derive Filler Reclassify Anonymize Binning RFM Analysis Ensemble Partition SetToFlag Restructure Transpose History Field Reorder Reproject Time Intervals

Server: Local Server

122MB / 136MB

Handling missing values (3)



Handling missing values (4)

Type

Preview

Types | Format | Annotations

Read Values | Clear Values | Clear All Values

Field	Measurement	Values	Missing	Check	Role
A1	Nominal	"?",a,b	*	None	Input
A2	Continuous	[13.75,80....	*	None	Input
A3	Continuous	[0.0,28.0]		None	Input
A4	Nominal	"?",l,u,y	*	None	Input
A5	Nominal	"?",g,gg,p	*	None	Input
A6	Nominal	"?",aa,c,cc...	*	None	Input
A7	Nominal	"?",bb,dd,f...	*	None	Input
A8	Continuous	[0.0,28.5]		None	Input
A9	Flag	t/f		None	Input

View current fields | View unused field settings

OK | Cancel | Apply | Reset

Type

Preview

Types | Format | Annotations

Read Values | Clear Values | Clear All Values

Field	Measurement	Values	Missing	Check	Role
A9	Flag	t/f		None	Input
A10	Flag	t/f		None	Input
A11	Continuous	[0.0,67.0]		None	Input
A12	Flag	t/f		None	Input
A13	Nominal	g,p,s		None	Input
A14	Continuous	[0.0,2000.0]	*	None	Input
A15	Continuous	[0.0,1000...		None	Input
A16	Flag	"-"/"+"		None	Input

View current fields | View unused field settings

OK | Cancel | Apply | Reset

Handling missing values (5)

Credit Approval project* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Select

Credit Approval.xlsx Type 16 Fields

Select

Preview

Settings Annotations

Mode: Include Discard

Condition:

OK Cancel Apply Reset

Expression Builder

Enter expression here or select from lists below.

General Functions

Function	Return
is_integer(ITEM)	Boolean
is_real(ITEM)	Boolean
is_number(ITEM)	Boolean
is_string(ITEM)	Boolean
is_date(ITEM)	Boolean
is_time(ITEM)	Boolean
is_timestamp(ITEM)	Boolean
is_datetime(ITEM)	Boolean
to_integer(ITEM)	Integer
to_real(ITEM)	Real

Fields

Type	Field	Storage
	A1	String
	A2	Real
	A3	Real
	A4	String
	A5	String
	A6	String
	A7	String
	A8	Real
	A9	String
	A10	String

is_integer(ITEM)

Returns a value of true if ITEM type is an integer. Otherwise, returns a value of false.

Check expression before saving

OK Cancel Check Help

Streams Outputs Models

Stream1 Credit ... Stream3

Classes

Business Understanding Data Understanding Data Preparation Modeling Evaluation Deployment

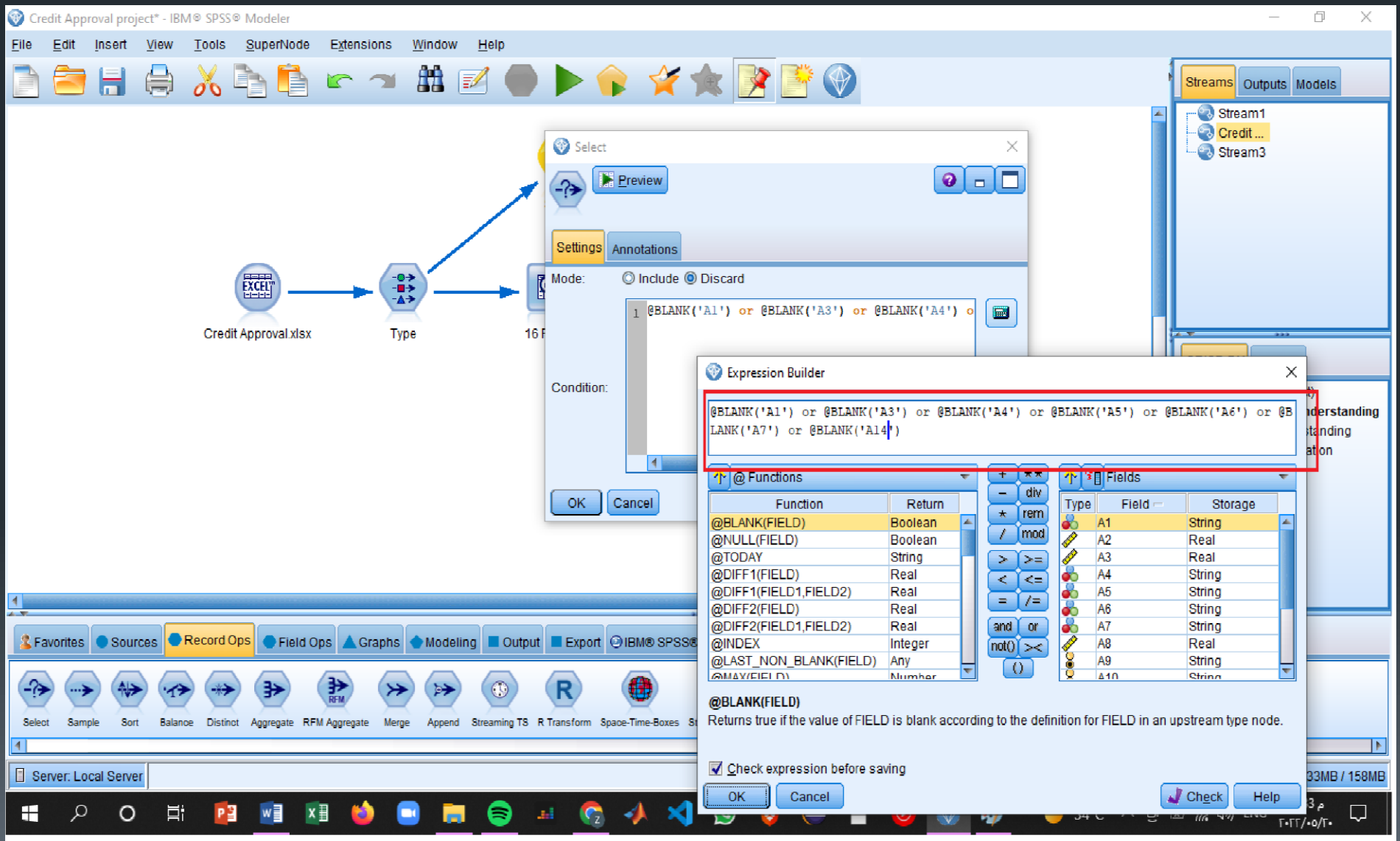
Export IBM® SPSS® Statistics

R Transform Space-Time-Boxes Streaming TCM

Server: Local Server

137MB / 143MB

Handling missing values (6)



Before Handling missing values

Credit Approval project* - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

Data Audit of [16 fields] #6

File Edit Generate

Audit Quality Annotations

Complete fields (%): 56.25% Complete records (%): 94.64%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value
A1	Nominal	--	--	--	Never	Fixed	98.261	678	0
A2	Continuous	9	0	None	Never	Fixed	98.261	678	12
A3	Continuous	10	0	None	Never	Fixed	100	690	0
A4	Nominal	--	--	--	Never	Fixed	99.13	684	0
A5	Nominal	--	--	--	Never	Fixed	99.13	684	0
A6	Nominal	--	--	--	Never	Fixed	98.696	681	0
A7	Nominal	--	--	--	Never	Fixed	98.696	681	0
A8	Continuous	19	3	None	Never	Fixed	100	690	0
A9	Flag	--	--	--	Never	Fixed	100	690	0
A10	Flag	--	--	--	Never	Fixed	100	690	0
A11	Continuous	6	2	None	Never	Fixed	100	690	0
A12	Flag	--	--	--	Never	Fixed	100	690	0
A13	Nominal	--	--	--	Never	Fixed	100	690	0
A14	Continuous	7	2	None	Never	Fixed	98.116	677	13
A15	Continuous	2	4	None	Never	Fixed	100	690	0
A16	Flag	--	--	--	Never	Fixed	100	690	0

OK

Server: Local Server

131MB / 158MB

After Handling missing values

Data Audit of [16 fields] #12

File Edit Generate

Audit Quality Annotations

Complete fields (%): 100% Complete records (%): 100%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space
A1	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A2	Continuous	7	0	None	Never	Fixed	100	653	0	0	
A3	Continuous	10	0	None	Never	Fixed	100	653	0	0	
A4	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A5	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A6	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A7	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A8	Continuous	18	3	None	Never	Fixed	100	653	0	0	
A9	Flag	--	--	--	Never	Fixed	100	653	0	0	
A10	Flag	--	--	--	Never	Fixed	100	653	0	0	
A11	Continuous	4	2	None	Never	Fixed	100	653	0	0	
A12	Flag	--	--	--	Never	Fixed	100	653	0	0	
A13	Nominal	--	--	--	Never	Fixed	100	653	0	0	
A14	Continuous	6	1	None	Never	Fixed	100	653	0	0	
A15	Continuous	1	4	None	Never	Fixed	100	653	0	0	
A16	Flag	--	--	--	Never	Fixed	100	653	0	0	

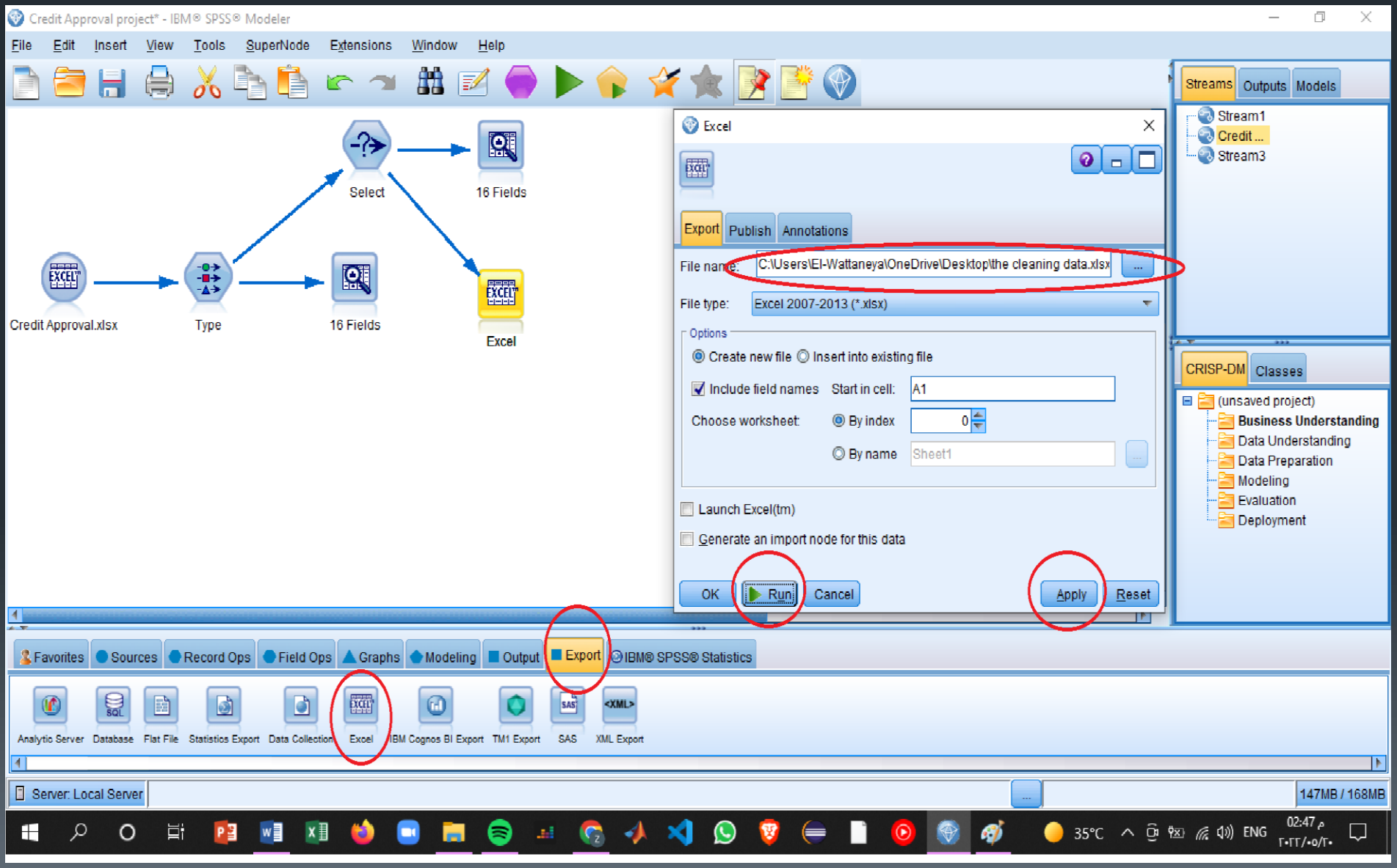
OK

Select Sample Sort Balance Distinct Aggregate RFM Aggregate Merge Append Streaming TS R Transform Space-Time-Boxes Streaming TCM

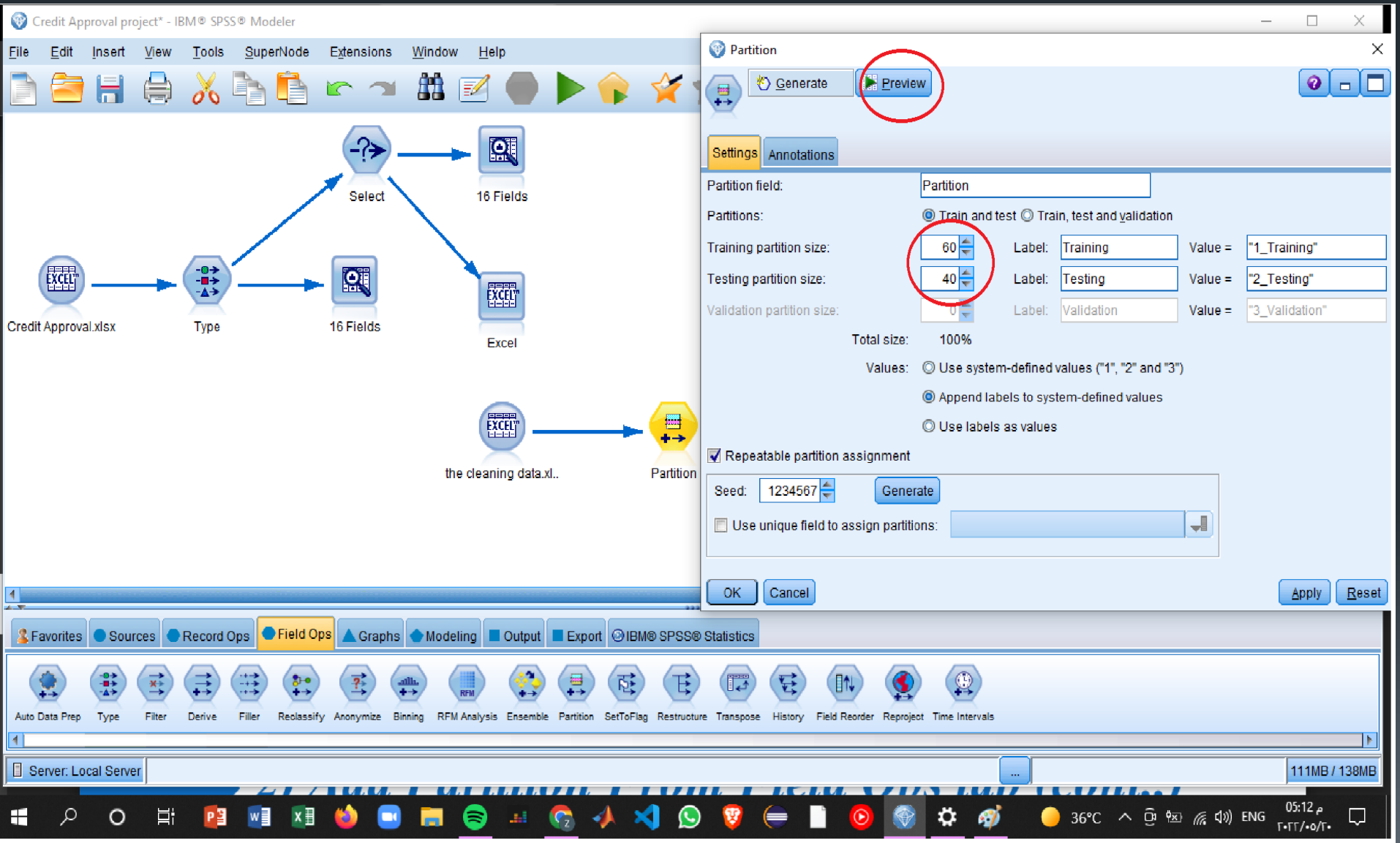
Server: Local Server

121MB / 168MB

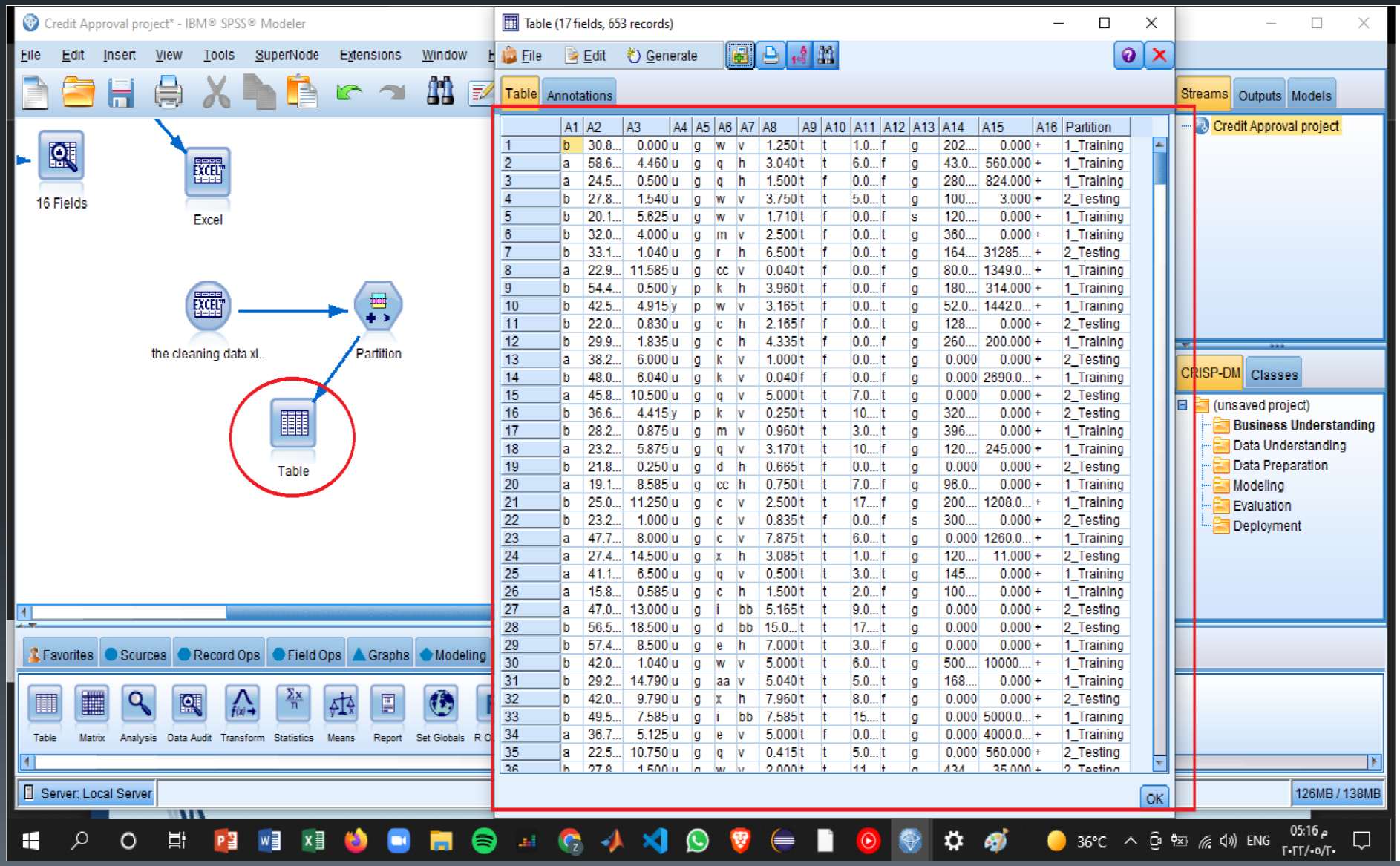
Classification (1)
Adding Handled Dataset



Classification (2)
Add Partitioning



Classification (3)
Appear Data after partitioning



	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	Partition
1	b	30.8...	0.000	u	g	w	v	1.250	t	t	1.0...	f	g	202....	0.000	+	1_Training
2	a	58.6...	4.460	u	g	q	h	3.040	t	t	6.0...	f	g	43.0...	560.000	+	1_Training
3	a	24.5...	0.500	u	g	q	h	1.500	t	f	0.0...	f	g	280....	824.000	+	1_Training
4	b	27.8...	1.540	u	g	w	v	3.750	t	t	5.0...	t	g	100....	3.000	+	2_Testing
5	b	20.1...	5.625	u	g	w	v	1.710	t	f	0.0...	f	s	120....	0.000	+	1_Training
6	b	32.0...	4.000	u	g	m	v	2.500	t	f	0.0...	t	g	360....	0.000	+	1_Training
7	b	33.1...	1.040	u	g	r	h	6.500	t	f	0.0...	t	g	164....	31285....	+	2_Testing
8	a	22.9...	11.585	u	g	cc	v	0.040	t	f	0.0...	f	g	80.0...	1349.0...	+	1_Training
9	b	54.4...	0.500	y	p	k	h	3.960	t	f	0.0...	f	g	180....	314.000	+	1_Training
10	b	42.5...	4.915	y	p	w	v	3.165	t	f	0.0...	t	g	52.0...	1442.0...	+	1_Training
11	b	22.0...	0.830	u	g	c	h	2.165	f	f	0.0...	t	g	128....	0.000	+	2_Testing
12	b	29.9...	1.835	u	g	c	h	4.335	t	f	0.0...	f	g	260....	200.000	+	1_Training
13	a	38.2...	6.000	u	g	k	v	1.000	t	f	0.0...	t	g	0.000	0.000	+	2_Testing
14	b	48.0...	6.040	u	g	k	v	0.040	f	f	0.0...	f	g	0.000	2690.0...	+	1_Training
15	a	45.8...	10.500	u	g	q	v	5.000	t	t	7.0...	t	g	0.000	0.000	+	2_Testing
16	b	36.6...	4.415	y	p	k	v	0.250	t	t	10....	t	g	320....	0.000	+	2_Testing
17	b	28.2...	0.875	u	g	m	v	0.960	t	t	3.0...	t	g	396....	0.000	+	1_Training
18	a	23.2...	5.875	u	g	q	v	3.170	t	t	10....	f	g	120....	245.000	+	1_Training
19	b	21.8...	0.250	u	g	d	h	0.665	t	f	0.0...	t	g	0.000	0.000	+	2_Testing
20	a	19.1...	8.585	u	g	cc	h	0.750	t	t	7.0...	f	g	96.0...	0.000	+	1_Training
21	b	25.0...	11.250	u	g	c	v	2.500	t	t	17....	f	g	200....	1208.0...	+	1_Training
22	b	23.2...	1.000	u	g	c	v	0.835	t	f	0.0...	f	s	300....	0.000	+	2_Testing
23	a	47.7...	8.000	u	g	c	v	7.875	t	t	6.0...	t	g	0.000	1260.0...	+	1_Training
24	a	27.4...	14.500	u	g	x	h	3.085	t	t	1.0...	f	g	120....	11.000	+	2_Testing
25	a	41.1...	6.500	u	g	q	v	0.500	t	t	3.0...	t	g	145....	0.000	+	1_Training
26	a	15.8...	0.585	u	g	c	h	1.500	t	t	2.0...	f	g	100....	0.000	+	1_Training
27	a	47.0...	13.000	u	g	i	bb	5.165	t	t	9.0...	t	g	0.000	0.000	+	2_Testing
28	b	56.5...	18.500	u	g	d	bb	15.0...	t	t	17....	t	g	0.000	0.000	+	2_Testing
29	b	57.4...	8.500	u	g	e	h	7.000	t	t	3.0...	f	g	0.000	0.000	+	1_Training
30	b	42.0...	1.040	u	g	w	v	5.000	t	t	6.0...	t	g	500....	10000....	+	1_Training
31	b	29.2...	14.790	u	g	aa	v	5.040	t	t	5.0...	t	g	168....	0.000	+	1_Training
32	b	42.0...	9.790	u	g	x	h	7.960	t	t	8.0...	f	g	0.000	0.000	+	2_Testing
33	b	49.5...	7.585	u	g	i	bb	7.585	t	t	15....	t	g	0.000	5000.0...	+	1_Training
34	a	36.7...	5.125	u	g	e	v	5.000	t	f	0.0...	t	g	0.000	4000.0...	+	1_Training
35	a	22.5...	10.750	u	g	q	v	0.415	t	t	5.0...	t	g	0.000	560.000	+	2_Testing
36	b	27.8...	1.500	u	g	w	v	2.000	t	t	11....	t	g	434....	35.000	+	2_Testing

Classification (4)

[A1]-> Record id, [A16] -> Target

Classification + Handling Missing Values - IBM® SPSS® Modeler

File Edit Insert View Tools SuperNode Extensions Window Help

handled Data.xlsx → Partition → Type

Partition → Table

Type

Preview

Types Format Annotations

Read Values | Clear Values | Clear All Values

Field	Measurement	Values	Missing	Check	Role
A1	Flag	b/a		None	Record ID
A2	Continuous	[13.75,76....		None	Input
A3	Continuous	[0.0,28.0]		None	Input
A4	Nominal	l,u,y		None	Input
A5	Nominal	g,gg,p		None	Input
A6	Nominal	aa,c,cc,d,e...		None	Input
A7	Nominal	bb,dd,ff,h,j,...		None	Input
A8	Continuous	[0.0,28.5]		None	Input
A9	Flag	t/f		None	Input
A10	Flag	t/f		None	Input
A11	Continuous	[0.0,67.0]		None	Input
A12	Flag	t/f		None	Input
A13	Nominal	g,p,s		None	Input
A14	Continuous	[0.0,2000.0]		None	Input
A15	Continuous	[0.0,10000...		None	Input
A16	Flag	"-"/"+"		None	Target
Partition	Nominal	"1_Trainin...		None	Partition

View current fields | View unused field settings

OK | Cancel | Apply | Reset

Streams | Outputs | Models

Classification + Handling Miss

CRISP-DM | Classes

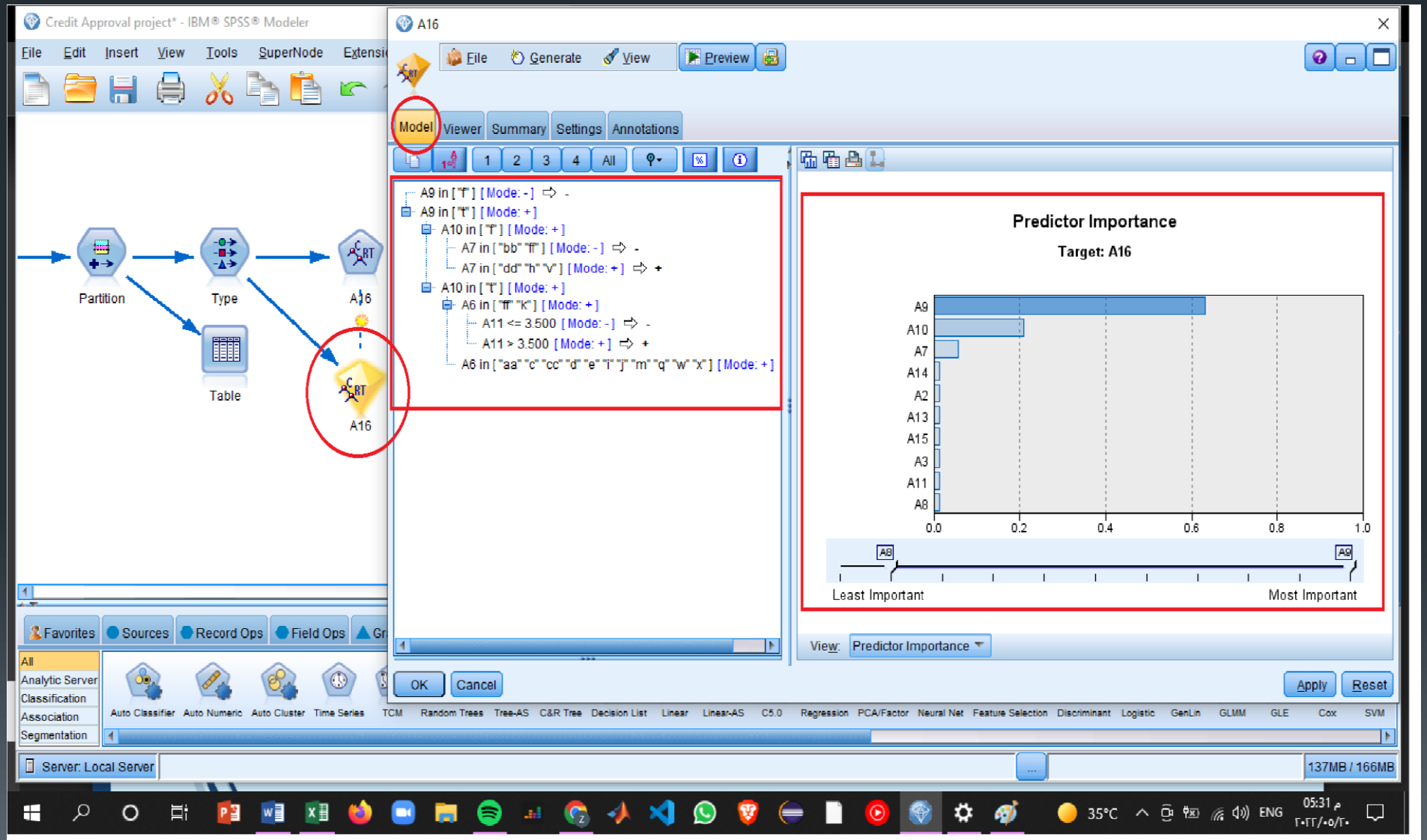
(unsaved project)
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment

Favorites | Sources | Record Ops | Field Ops | Graphs | Modeling | Output | Export | IBM® SPSS® Statistics

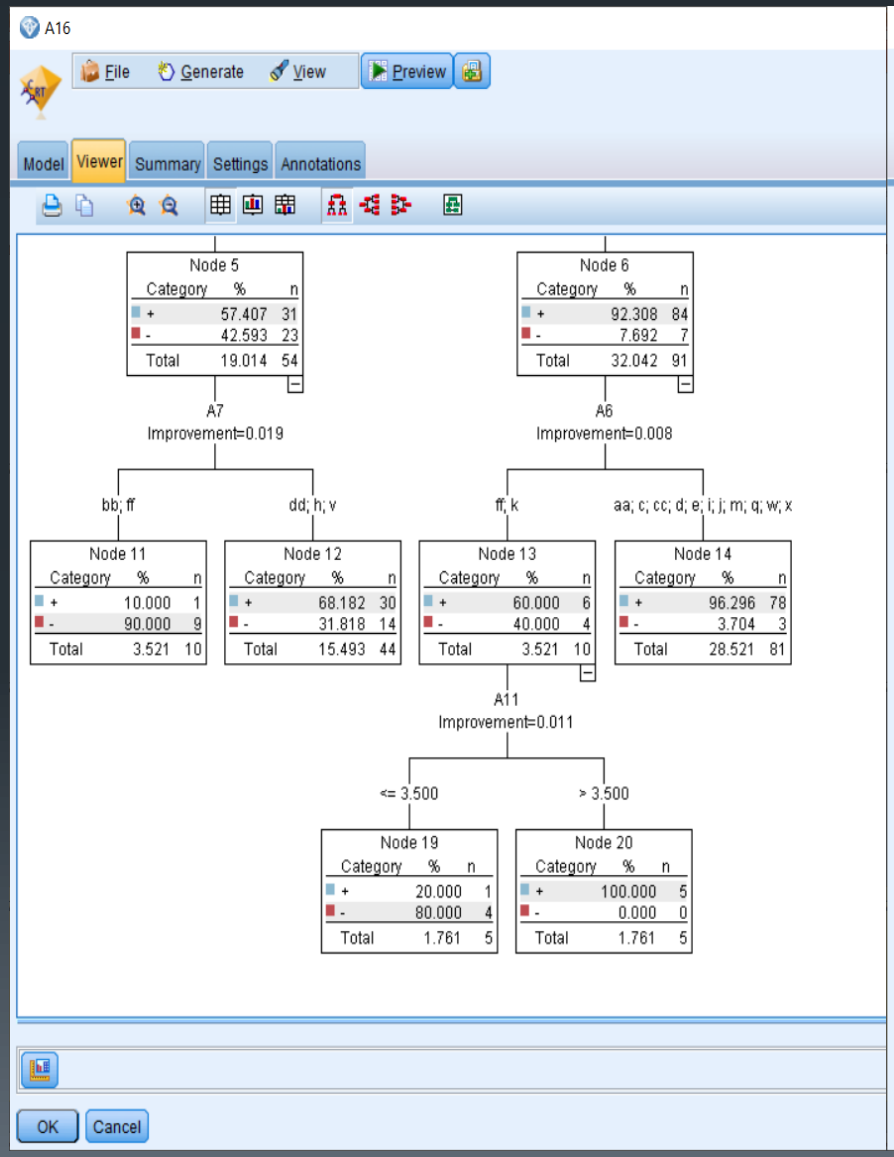
Database | Var. File | Auto Data Prep | Select | Sample | Aggregate | Derive | Type | Filter | Graphboard | Auto Classifier | Auto Numeric | Auto Cluster | Table | Flat File | Database

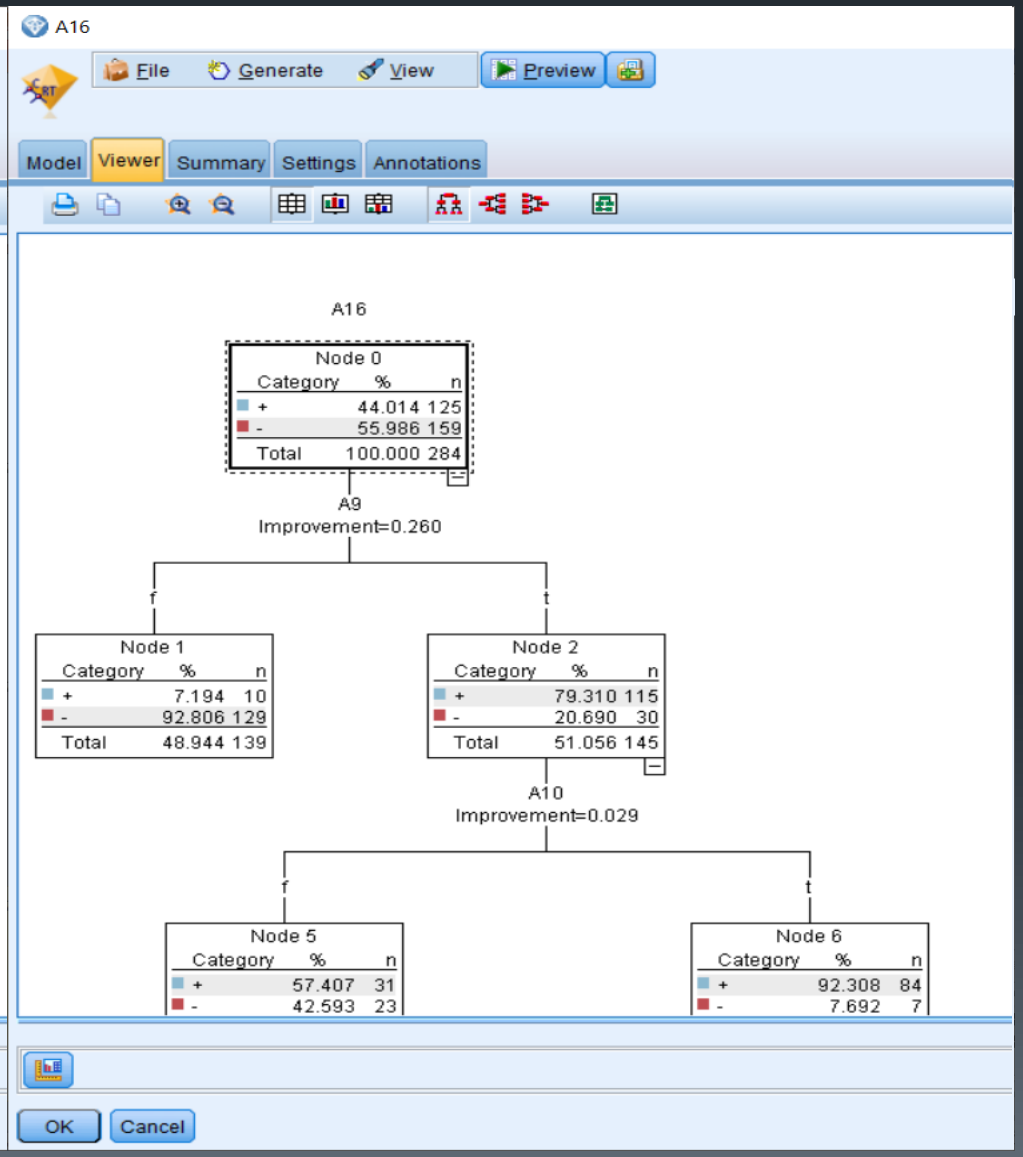
Server: Local Server | 108MB / 150MB

Classification (5)
Modeling Dataset

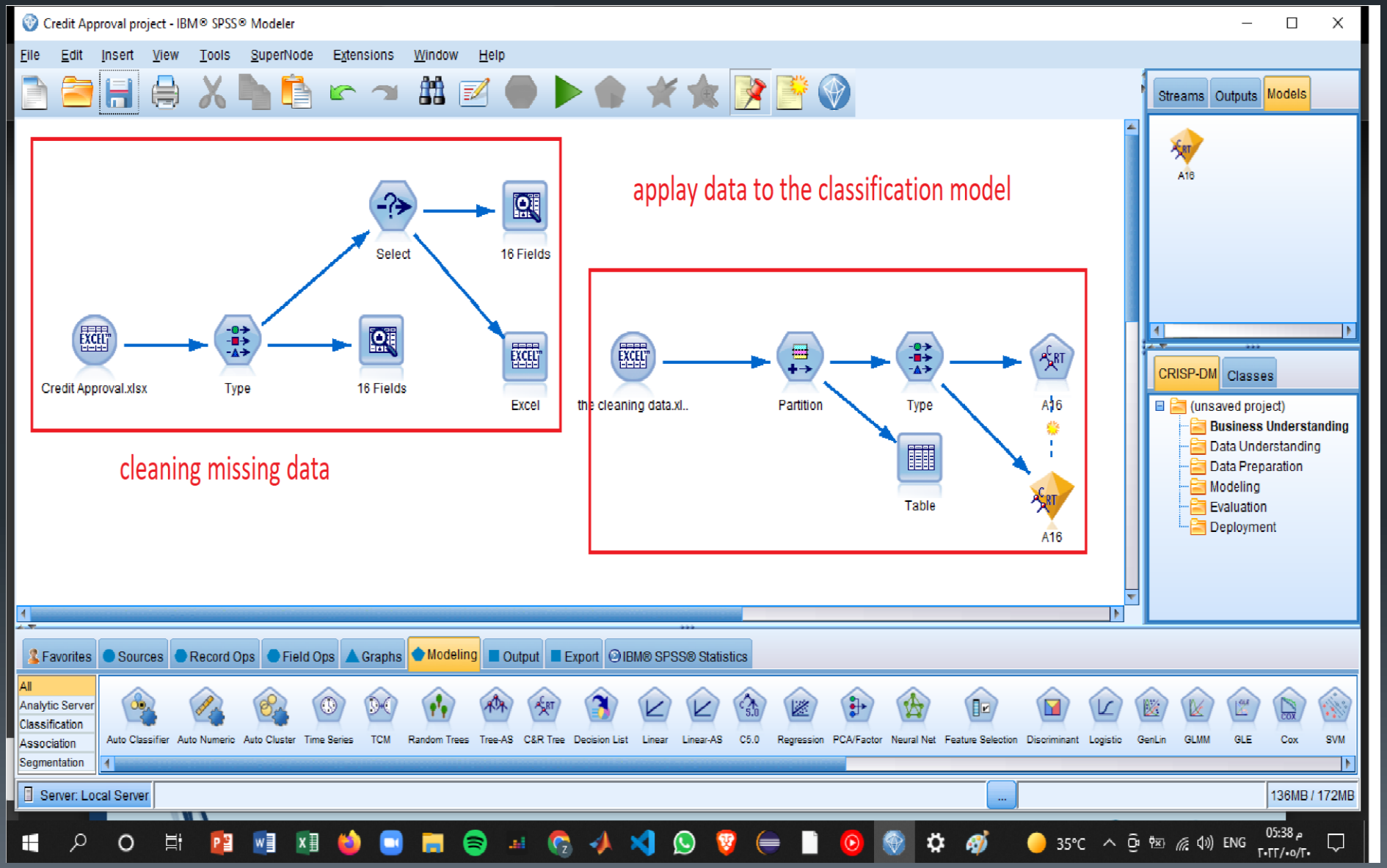


Classification (6)
Viewing Dataset

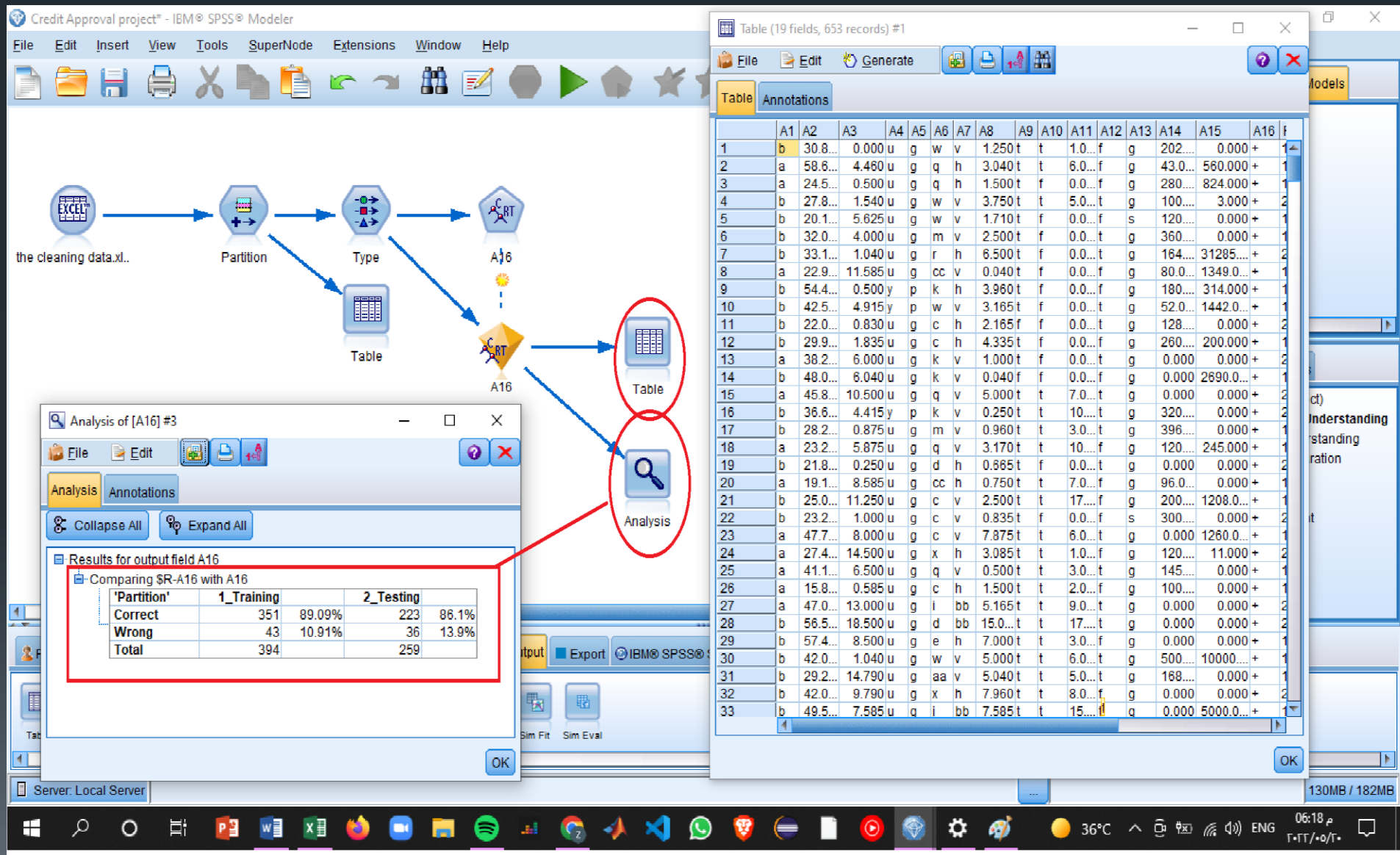




Cleaning Missing Values + Classification



Performance Evaluation for the Classification Model



THANKS!