

Watson Studio SPSS Modeler Overview

Overview

In this lab you will learn how to implement analytics in **SPSS Modeler**, a well-known visual data mining workbench which is part of **Watson Studio**. The lab will introduce the SPSS Modeler capability using the Titanic dataset. The lab will guide the development of an SPSS Modeler stream that will prepare the input data to train and evaluate a machine learning model for predicting survivability of a passenger on the Titanic.







Introduction

SPSS Modeler is a visual data mining workbench. Modeler can be used to complete all tasks in analytic application development

- Data understanding
- Data preparation
- Model building
- Model evaluation

Assets developed in Modeler are called “flows”. Another frequently used term in Modeler documentation is “streams” (used in Modeler desktop documentation). A flow starts with one or several data sources. Using visual nodes, a user can apply different operations to data. Data “flows” from one node to another in the direction of the arrows.

Visual nodes in modeler are color-coded and organized by type of operation: **Record Operations**, **Field Operations**, **Graphs**, **Modeling**, **Output**, and **Export** (data sources). Most operations are well-known functions in data preparation and analytics, such as sampling, filtering, binning, etc.

The data sources are purple	
Data preparation operations are blue	
Algorithms are green	
The models that are created based on algorithms are orange	
Different types of output (graphs, tables, external files) are black	
The nodes with a star icon are called “supernodes” because they contain several	

nodes. Supernodes are used for visual organization of the flow.

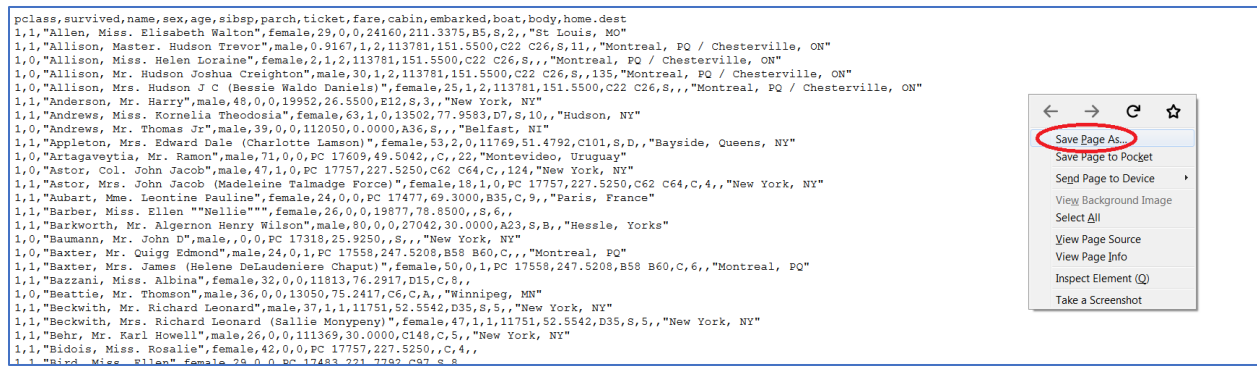
If a user needs more information about a particular node, it can be looked up in Modeler documentation. SPSS also publishes the **Algorithms Guide** that explains how machine learning algorithms are implemented in Modeler.

Lab Steps

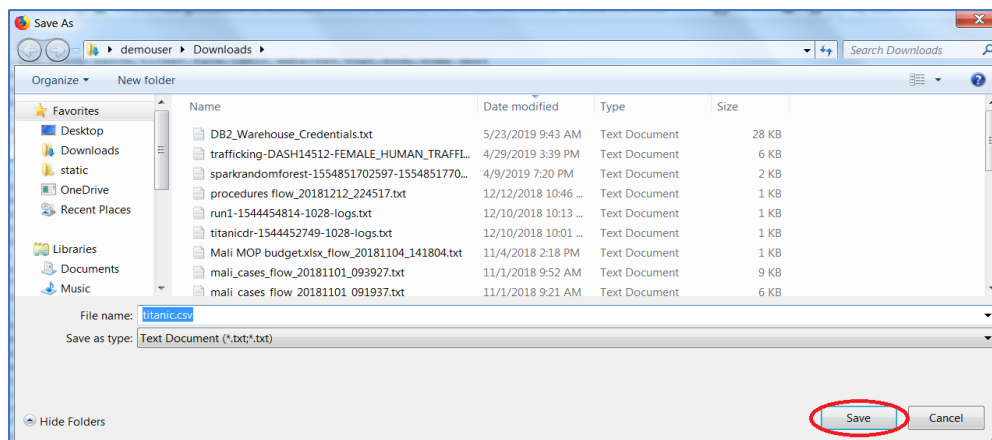
Step 1: Adding a Data Asset to the Watson Studio Labs project

This step can be skipped if the titanic.csv file was already downloaded in a previous lab.

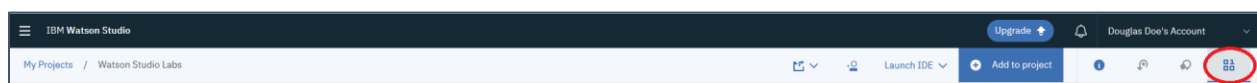
1. Download the Titanic data file from the following location by clicking [here](#).
2. Right-click on the screen and click on Save Page As ...



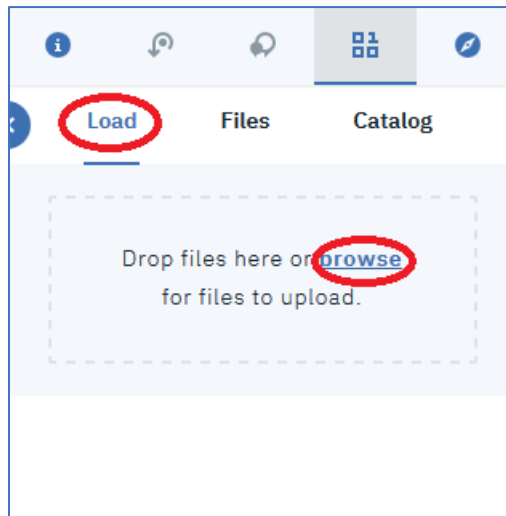
3. Click on **Save** to save the titanic.csv file.



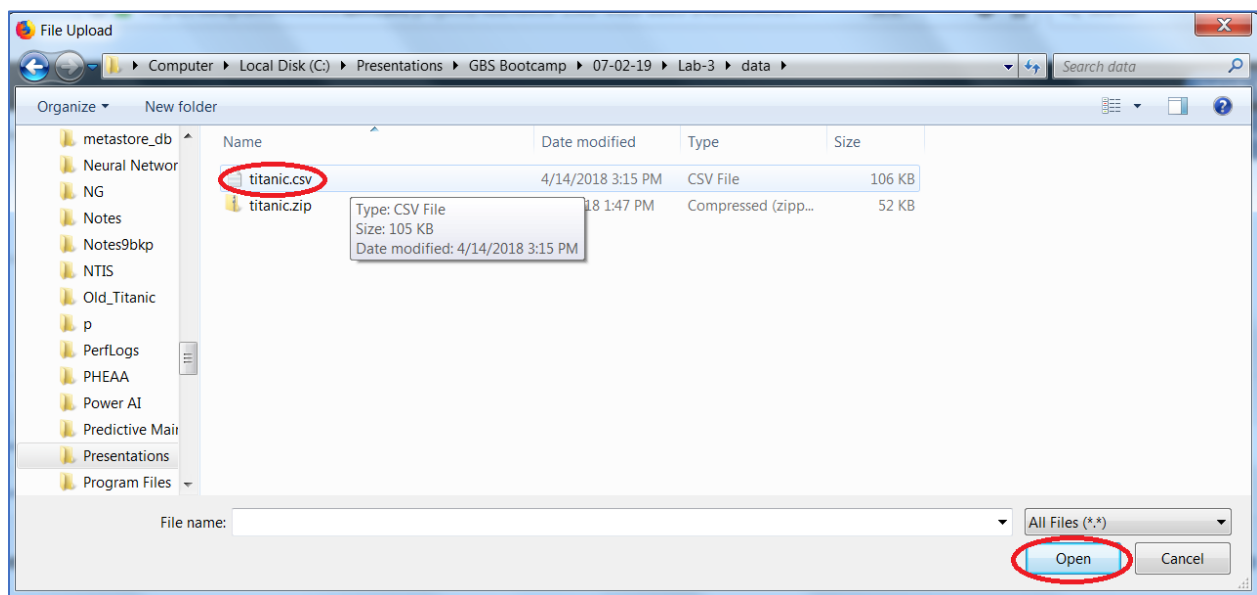
4. Go back to your Watson Studio Labs project. Click on the  icon.



5. Click on the **Load** tab and then click on **browse**. If you don't see the **Load** tab, click on the  icon again.



6. Go to the folder where the `titanic_csv` file is stored. Select the `titanic.csv` file and then click **Open**.



7. The file is now added as a Data Asset.

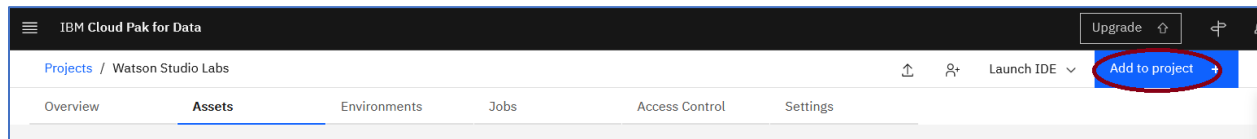


Step 2: Create a Model to predict survival

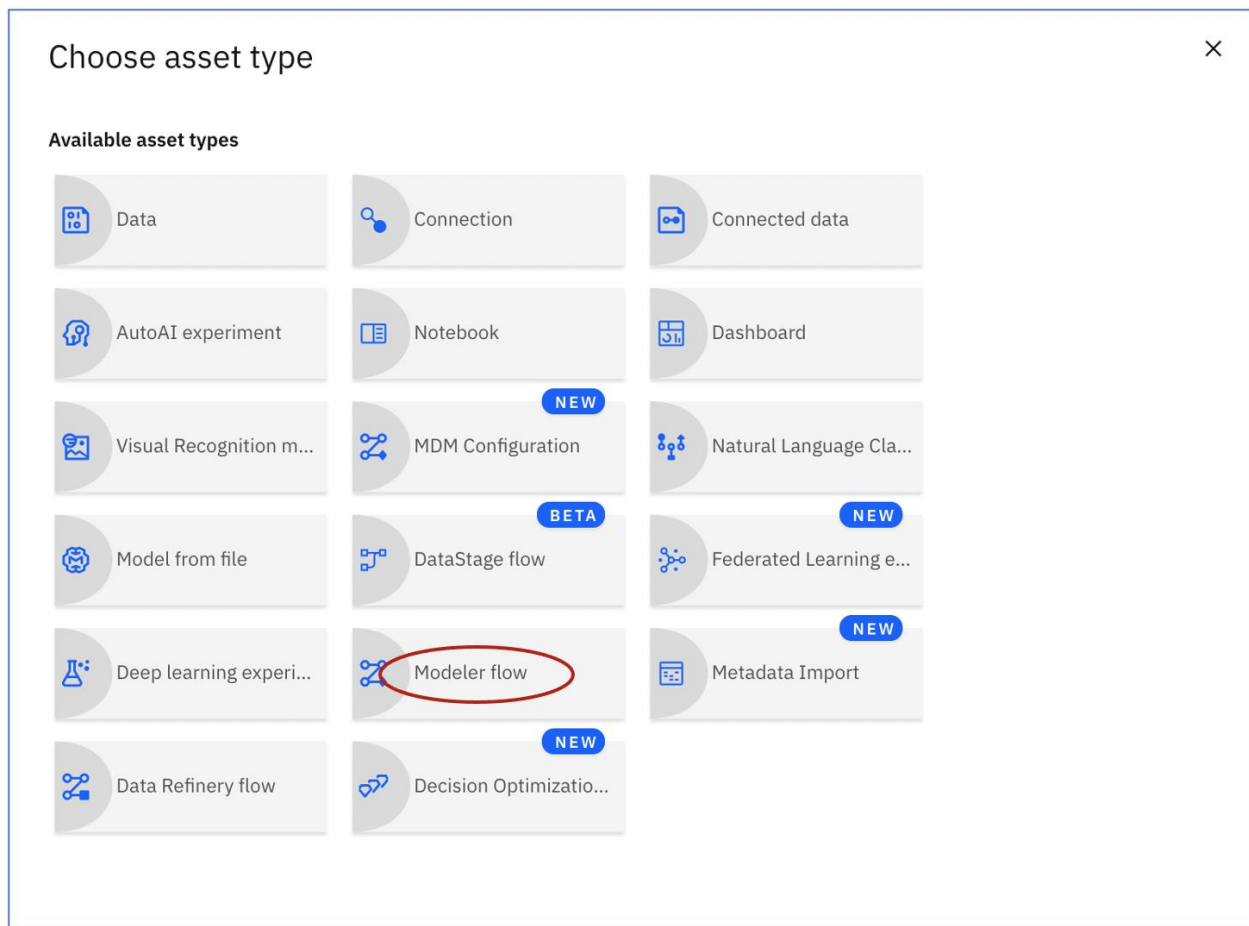
In this section, we will create a Machine Learning flow using SPSS nodes.

Step 2.1 Create a New Flow and Load the Data

1. In the Watson Studio project, click on **Add to project**.



2. Click on **Modeler flow**.



3. Enter a **Name** for the flow, optionally enter a **Description**, and click on **Create**.

New From File From Example

Name
Titanic SPSS

Description (optional)
Type description here.

Runtime
IBM SPSS Modeler

Cancel Create

4. A status message will appear. Please wait until the Flow Editor opens.

Creating a new Modeler flow

Status: Starting a runtime for the flow...
 The runtime has 4 vCPU + 16 GB RAM and consumes 2 capacity units per hour.
 This may take a minute.

5. Click on **Import** and then **Data Asset** and hold the left mouse key on the Data Asset icon and **drag it onto the left side of the canvas**. Release the left mouse key.

Projects / Watson Studio Labs / Titanic SPSS

Find palette nodes

Run Copy Paste Delete Zoom In Zoom Out

Import

Data Asset

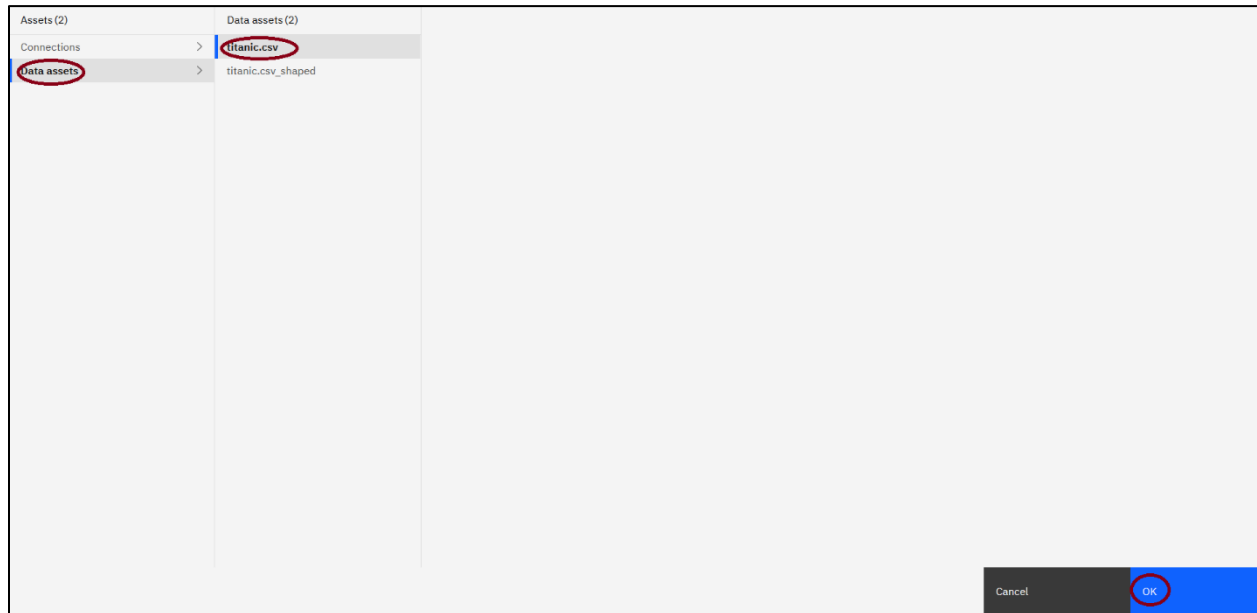
User Input

Sim Gen

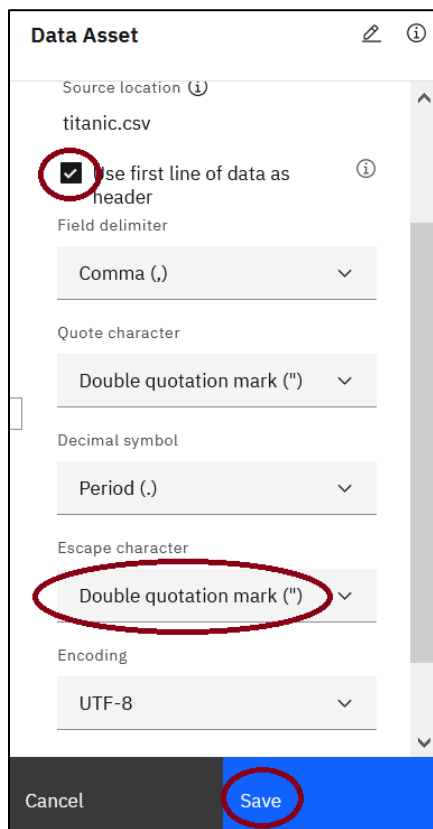
Extension Import


Data Asset

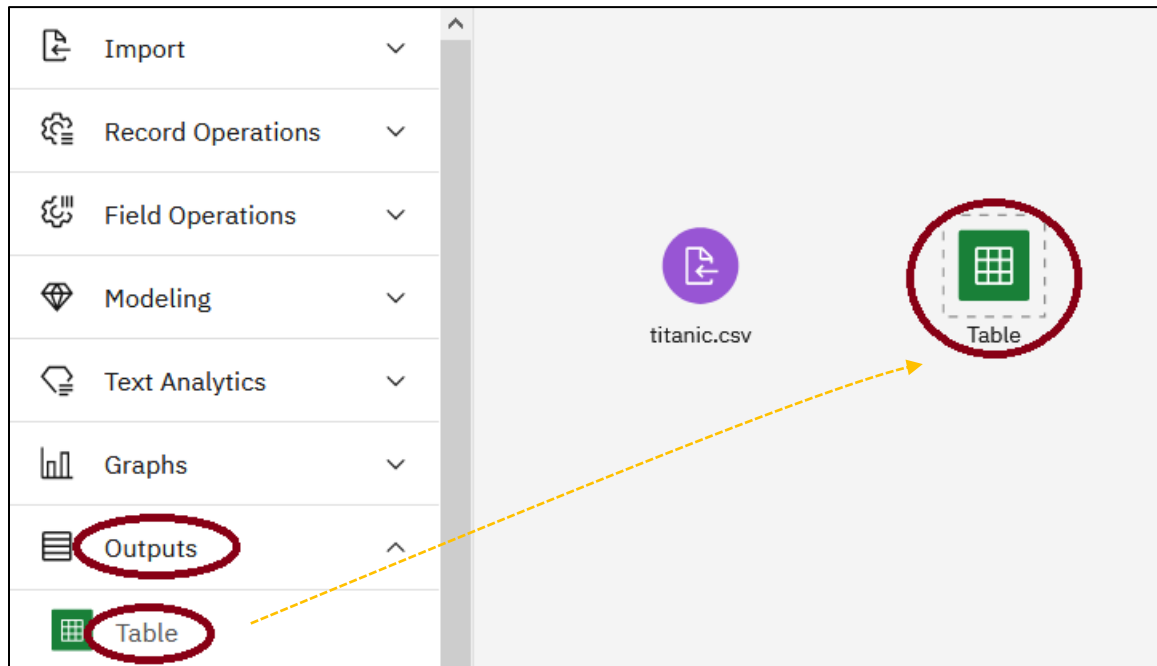
6. Double click on the **Data Asset**. Click on **Data assets**, **titanic.csv** and click **OK**.




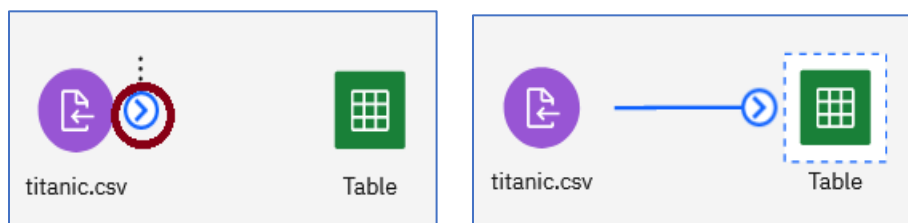
7. Click on the **Use first time of data as header** check box and set **Escape character** to **Double quotation mark (")**. Click **Save**.



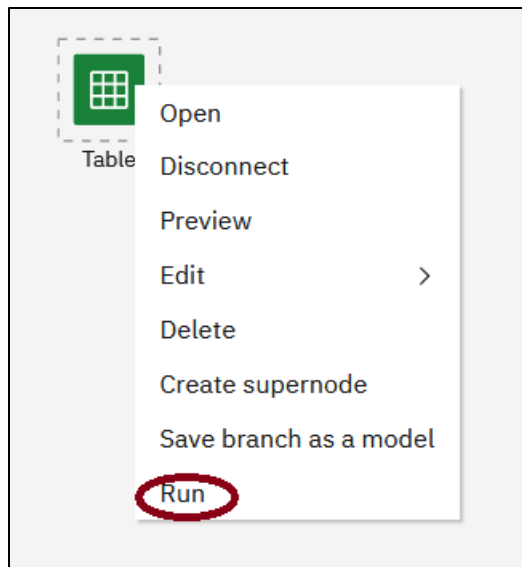
8. Click on the **Outputs** menu item in the Node Palette on the left and then click on the **Table** icon and drag the icon to the right of the titanic.csv icon. The SPSS Table node will display the contents of the csv file. If the Node Palette is not visible, click on the Node Palette icon .



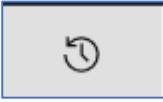

9. Connect the right side of the titanic.csv icon to the left side of the Table icon. This is accomplished by hovering the mouse over the titanic.csv icon, clicking on the arrow  at the right side of the titanic.csv icon, holding the left mouse key and dragging the arrow until Table icon becomes active (dashed lines around it), and then releasing the left mouse key.

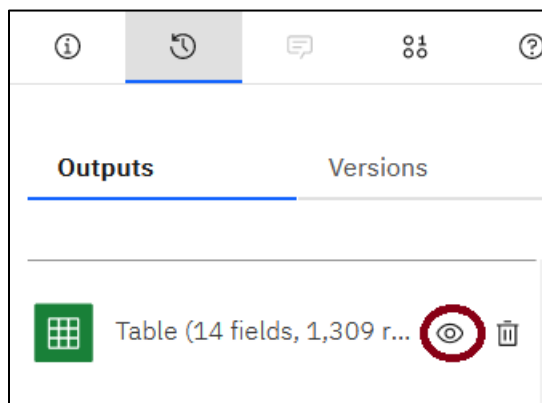


10. Right click on the **Table** icon and select **Run**.



11. The “Running Flow” prompt will appear and then when completed a Table output selection will appear on the right side of the screen under the **Outputs** tab. If the Table

output selection does not appear, select the  icon. Click on the  to view the titanic.csv contents.



12. Each row contains information on a passenger on the Titanic. We will use this data to make predictions on survivability.

PCLASS	SURVIVED	NAME	SEX	AGE	SIBSP	PARCH	TICKET	FARE	CABIN	EMBARKED	BOAT
1	1	Allen, Miss. Elisabeth	female	29	0	0	24100	211.3375	B5	S	2
1	1	Allison, Master. Huds	male	0.9167	1	2	113781	151.55	C22 C26	S	11
1	0	Allison, Miss. Helen L	female	2	1	2	113781	151.55	C22 C26	S	
1	0	Allison, Mr. Hudson J	male	30	1	2	113781	151.55	C22 C26	S	
1	0	Allison, Mrs. Hudson	female	25	1	2	113781	151.55	C22 C26	S	
1	1	Anderson, Mr. Harry	male	48	0	0	19952	26.55	E12	S	3
1	1	Andrews, Miss. Korn	female	63	1	0	13502	77.9583	D7	S	10
1	0	Andrews, Mr. Thoma	male	39	0	0	112050	0	A36	S	
1	1	Appleton, Mrs. Edwa	female	53	2	0	11769	51.4792	C101	S	0
1	0	Artagaveytia, Mr. Ra	male	71	0	0	PC 17609	49.5042		C	
1	0	Astor, Col. John Jaco	male	47	1	0	PC 17757	227.525	C82 C84	C	
1	1	Astor, Mrs. John Jaco	female	18	1	0	PC 17757	227.525	C82 C84	C	4
1	1	Aubart, Mme. Leonie	female	24	0	0	PC 17477	69.3	B35	C	9
1	1	Barber, Miss. Ellen ?	female	26	0	0	19877	78.85		S	6
1	1	Barkworth, Mr. Alger	male	80	0	0	27042	30	A23	S	8


Page 1 / 7

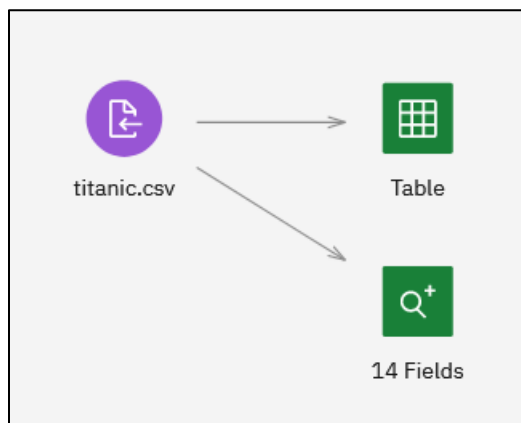
13. Return to the SPSS canvas by clicking on Titanic SPSS (or what you named the flow) in the breadcrumb area.

My Projects / Watson Studio Labs / Titanic SPSS / Table (14 fields, 1,309 records)					
pclass	survived	name	sex	age	sibsp

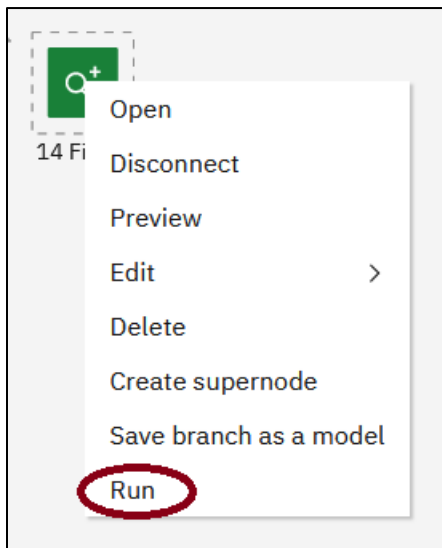
Step 2.2 Explore the Data using the Data Audit Node

Perusing through the data in the table, we can see that there are missing values. The SPSS Modeler has a Data Audit node that provides profiling information on the input data that is useful for cleansing the data. It provides a comprehensive first look at the data, including summary statistics, as well as information about outliers, missing values, and extremes.


1. Add a **Data Audit** node to the flow by clicking on the **Outputs** menu item in the Node Palette, and then dragging the **Data Audit** node to underneath the Table node. If the Node Palette is not visible, click on the Node Palette icon . Connect the titanic.csv node to the Data Audit node. The canvas should appear as below.

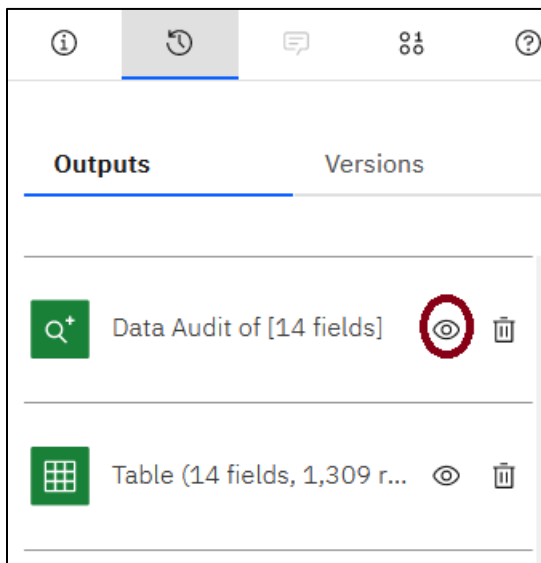


2. Right click on the **Data Audit** node and click **Run**.



3. The “Running Flow” prompt will appear and then when completed a Data Audit output selection will appear on the right side of the screen under the **Outputs** tab. If the **Outputs**

tab doesn't display, click on the  icon. Click on the  icon.



4. We can see that several fields have many missing values (cabin, boat, body, home_dest. These fields will be removed using a **Filter** node below. Other fields have only a few missing values (fare, embarked). Age has over 250 missing values. The rows containing the missing values for fare, embarked, and age will be removed using a **Select** node below.

My Projects / Watson Studio Labs / Titanic SPSS Flow / Data Audit of [14 fields]


</

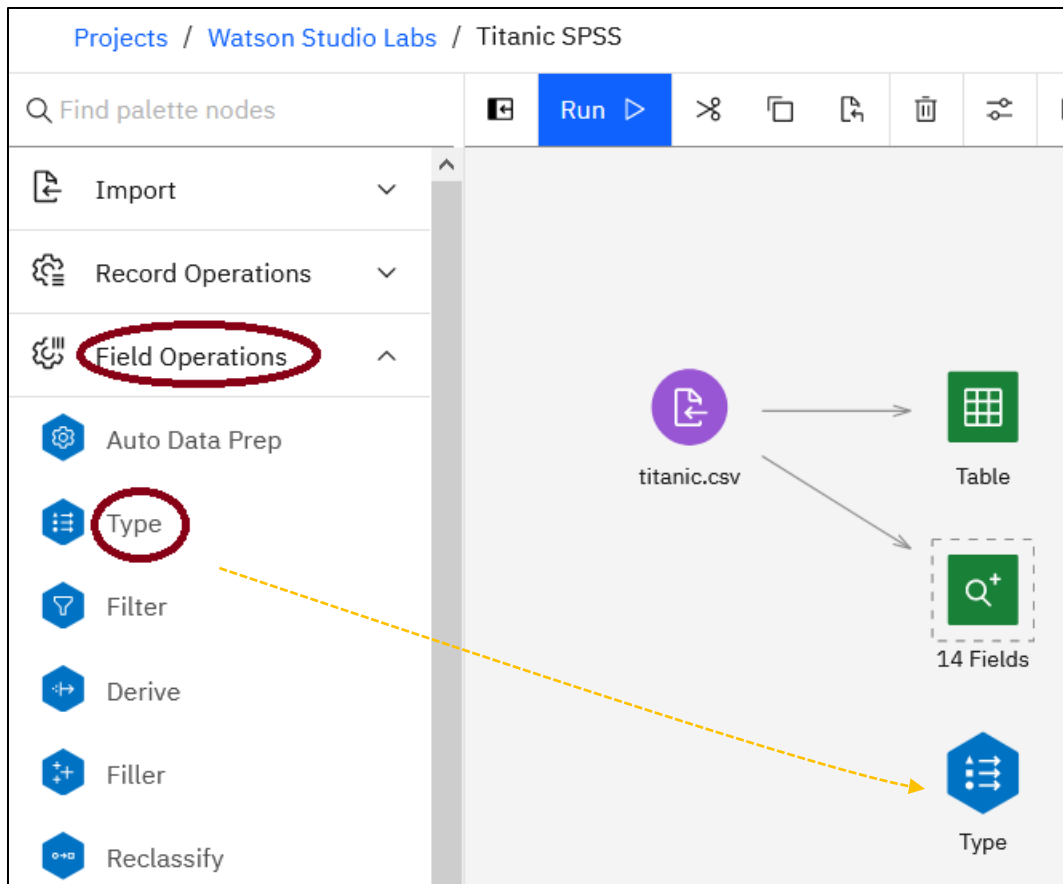
- Return to the SPSS canvas by clicking on the Titanic SPSS (or what you named the flow) in the breadcrumb area.

My Projects / Watson Studio Labs / Titanic SPSS / Data Audit of [14 fields]									
---	--	--	--	--	--	--	--	--	--

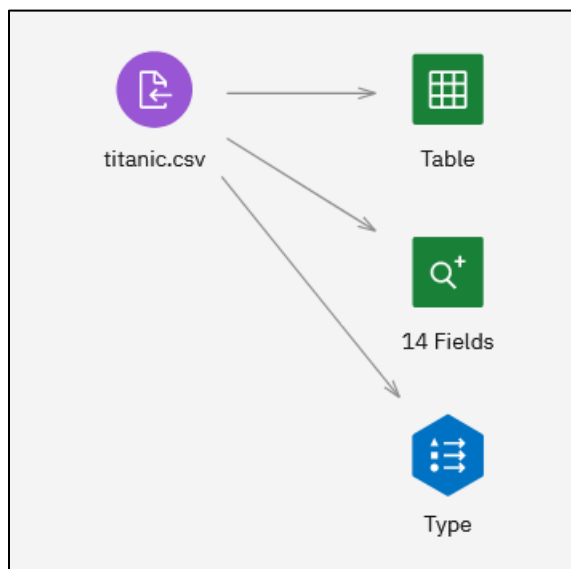
Step 2.3 Explore the Data using Graph Nodes.

Let's explore the data using Graph Nodes. The Distribution node, and the Histogram node will be used to explore some of the characteristics of the Titanic Data Set. First, we will add a Type node to the canvas. The Type node specifies field metadata and properties. We will change the measurement property for the "pclass" and "survived" fields that was derived as "Continuous" (by scanning the data values) to "Ordinal" and "Flag" respectively.

- Add a **Type** node to the flow by clicking on the **Field Operations** menu item in the Node Palette and then drag the **Type** node underneath the **Data Audit** node. If the Node Palette is not visible, click on the Node Palette icon .



2. Connect the titanic.csv node to the **Type** node. The canvas should appear as below.



3. Double click on the **Type** node. This will open a **Type** menu pallet on the right side of the screen.
4. Click on **Read Values**.

Type

Settings

Default Mode ⓘ
☒ Read metadata ☐ Pass (do not scan)

Type Operations

Read Values

Clear All Values

Find in column Field

<input type="checkbox"/>	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# pclass	Continuous ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	# survived	Continuous ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	abc name	Categorical ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	abc sex	Categorical ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	# age	Continuous ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	# sibsp	Continuous ▾	Input ▾	Read ▾		None ▾ ⚙
<input type="checkbox"/>	# parch	Continuous ▾	Input ▾	Read ▾		None ▾ ⚙

5. Certain fields are changed to a Typeless measure due to the number of unique values in the field. If there are a large number of unique values, the field is not useful for modeling.

Messages
X

Last run was 5 seconds ago

!
Node: Type

Large set type field 'name' has changed to typeless

!
Node: Type

Large set type field 'ticket' has changed to typeless

!
Node: Type

Large set type field 'home_dest' has changed to typeless

Clear all

- Select the dropdown in the **Measure** column next to **Survived**. Change the **Measure** from **Continuous** to **Flag**. Select the dropdown in the **Measure** column next to **pclass**. Change the measure from **Continuous** to **Ordinal**. Click **Save**.

Settings

Default Mode ⓘ

☒ Read metadata
 ☐ Pass (do not scan)

Type Operations

Read Values

Clear All Values

Find in column Field


	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# pclass	Ordinal	Input	Instantiated	1, 3	None
<input type="checkbox"/>	# survived	Flag	Input	Instantiated	0, 1	None
<input type="checkbox"/>	abc name	Typeless	None	...		None
<input type="checkbox"/>	abc sex	Flag	Input	Instantiated	female, male	None
<input type="checkbox"/>	# age	Continuous	Input	Instantiated	0.1667, 80.0	None
<input type="checkbox"/>	# sibsp	Continuous	Input	Instantiated	0, 8	None
<input type="checkbox"/>	# parch	Continuous	Input	Instantiated	0, 9	None
<input type="checkbox"/>	abc ticket	Typeless	None	...		None

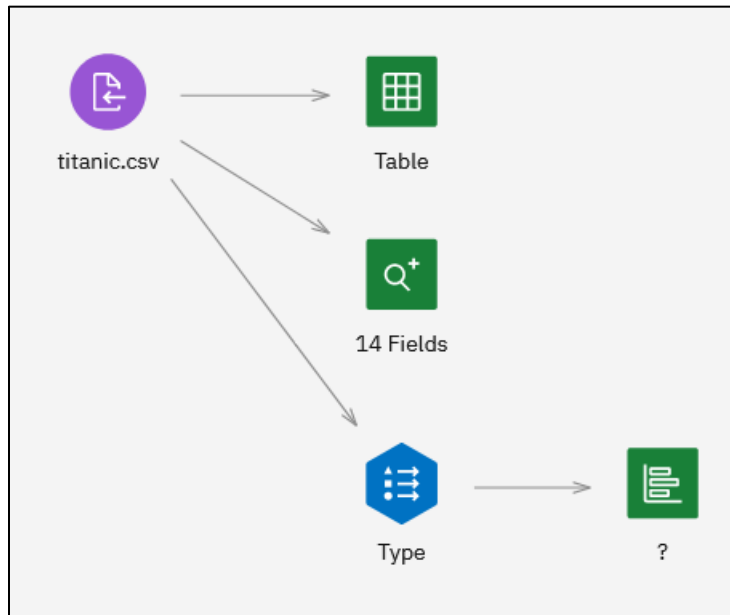
Format

Annotations

Cancel

Save

7. Add a **Distribution** node to the flow by clicking on the **Graph** menu item and then dragging the **Distribution** node to the canvas to the right of the **Type** node. If the Node Palette is not visible, click on the Node Palette icon . Connect the **Type** node to the **Distribution** node. The canvas should appear as below. The ? indicates that the fields to be plotted have not been identified.



8. Double click on the Distribution Node. Click on the **Plot** dropdown. In the **Field (discrete)** dropdown, select **pclass**. In the Color (discrete) dropdown, select **survived**. Click on the **normalize by color** checkbox, and then click **Save**.

Plot

Plot ⓘ

☒ Selected fields

☐ All flags (true values)

Field (discrete) ⓘ

pclass

▼

Color (discrete) ⓘ

survived

▼

☒

 Normalize by color

Sort ⓘ

☒ Alphabetic

☐ By count

☐ Proportional scale

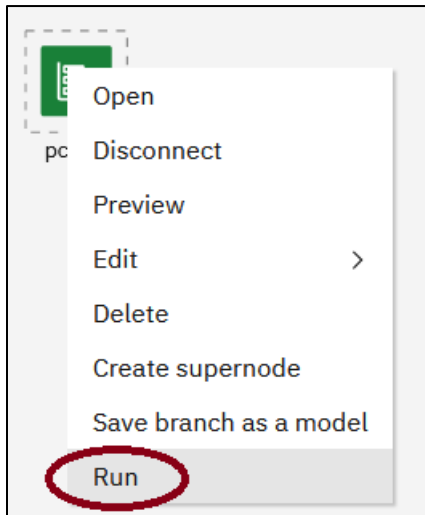
Appearance ▼


Annotations ▼

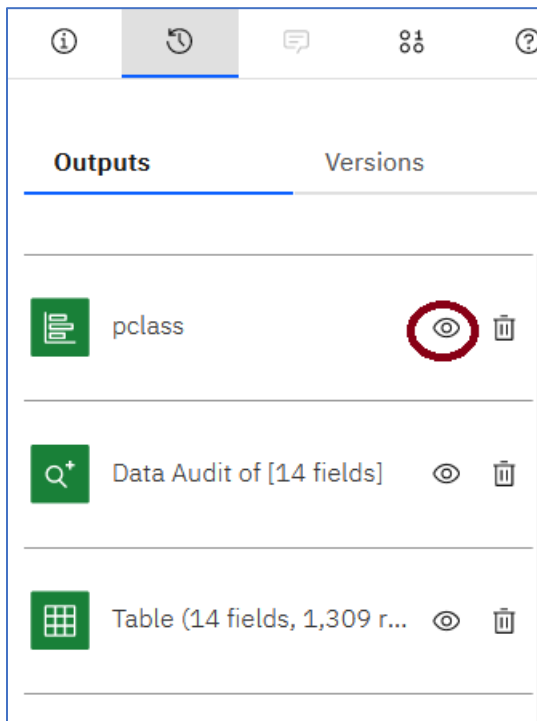
Cancel

Save

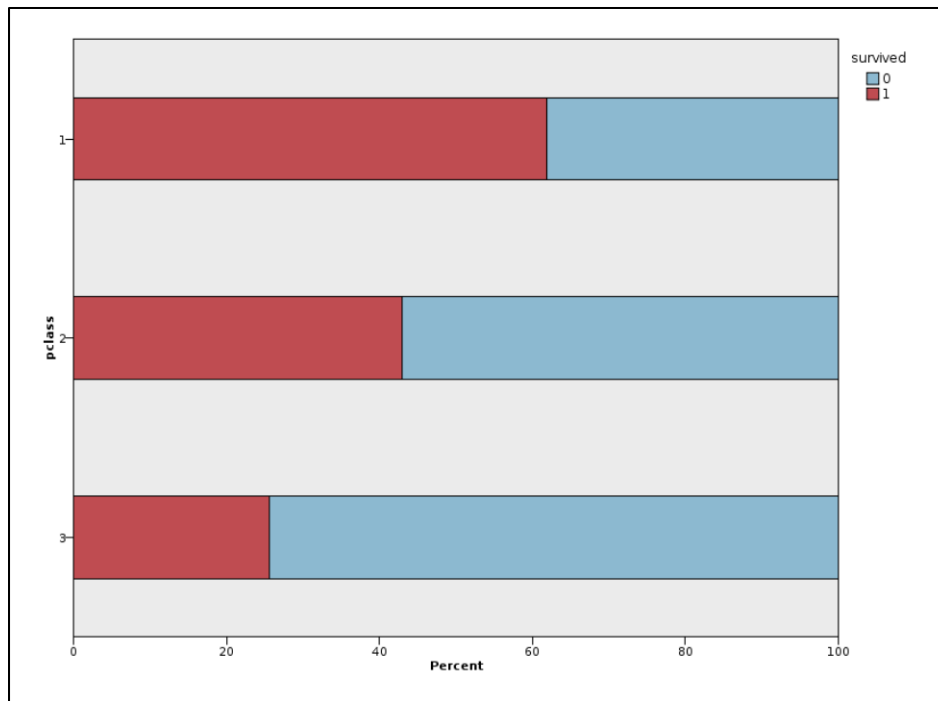
9. Right click on the Distribution node and select **Run**.




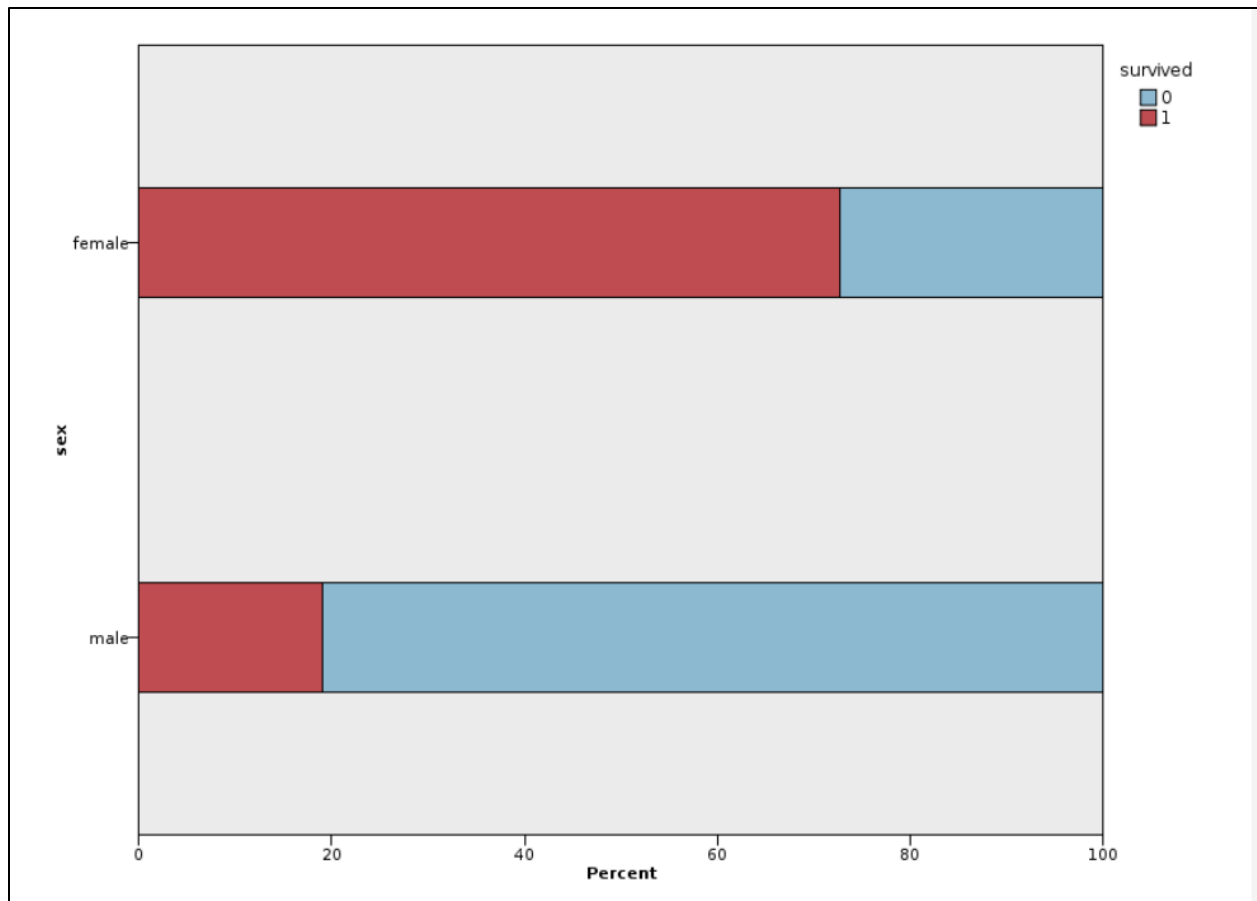
10. The Distribution of pclass output will appear under the **Outputs** tab. Click on the  icon.




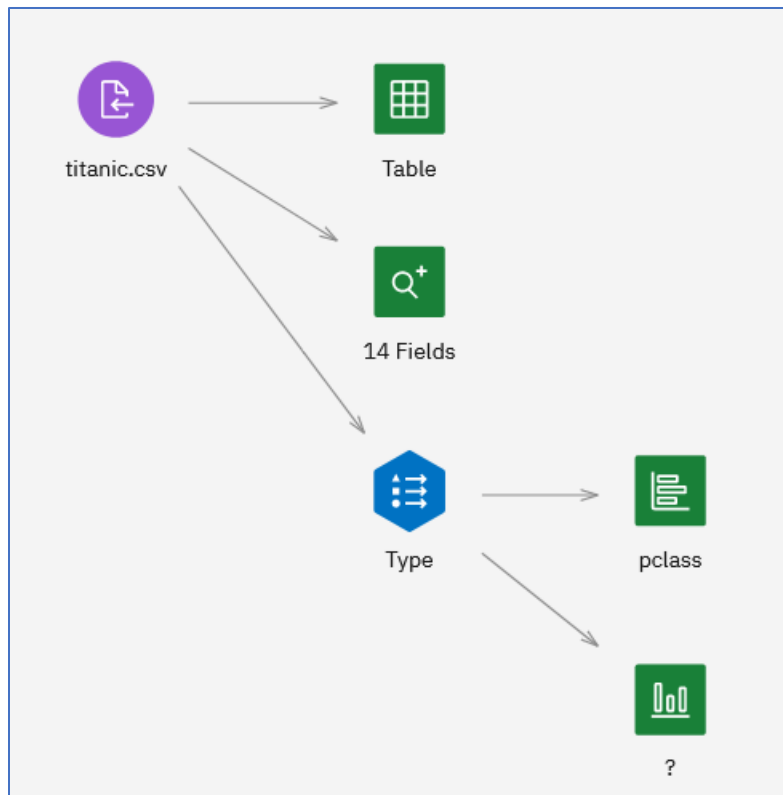
11. We can see from the graph that the likelihood of surviving is correlated to the passenger class. The first-class passengers have the highest rate of survivability. **Note if you see a graph with green bars, instead of the one below, redo Steps 9-11.**



12. Return to the SPSS canvas by clicking on **Titanic SPSS** in the breadcrumb area. You can change the distribution graph to show the survivability by gender by double clicking on the Distribution node and replacing **pclass** with **sex** and clicking Save. Re-run the graph by right clicking on the Distribution node and selecting Run. Click on the  next to **sex** to display the graph.



13. Return to the SPSS canvas by clicking on **Titanic SPSS** in the breadcrumb area. Add a **Histogram** node to the flow by clicking on the **Graphs** menu item and then dragging the **Histogram** node to the canvas underneath the **Distribution** node. If the Node Palette is not visible, click on the Node Palette icon . Connect the **Type** node to the **Histogram** node. The canvas should appear as below. The ? indicates that the fields to be plotted have not been identified.



14. Double click on the **Histogram** node. Click on the **Plot** dropdown. Select **fare** from the Field (continuous) dropdown. Select **survived** from the Color (discrete) dropdown. Click on **Save**.

Plot

Field (continuous) ⓘ
fare

Color (discrete) ⓘ
survived

Panel (discrete) ⓘ
...

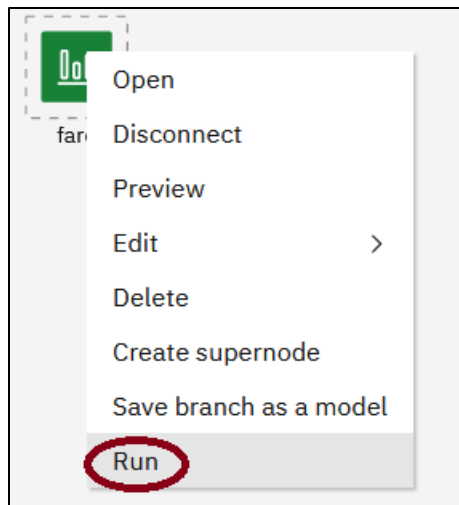
Options


Appearance

Annotations

CancelSave

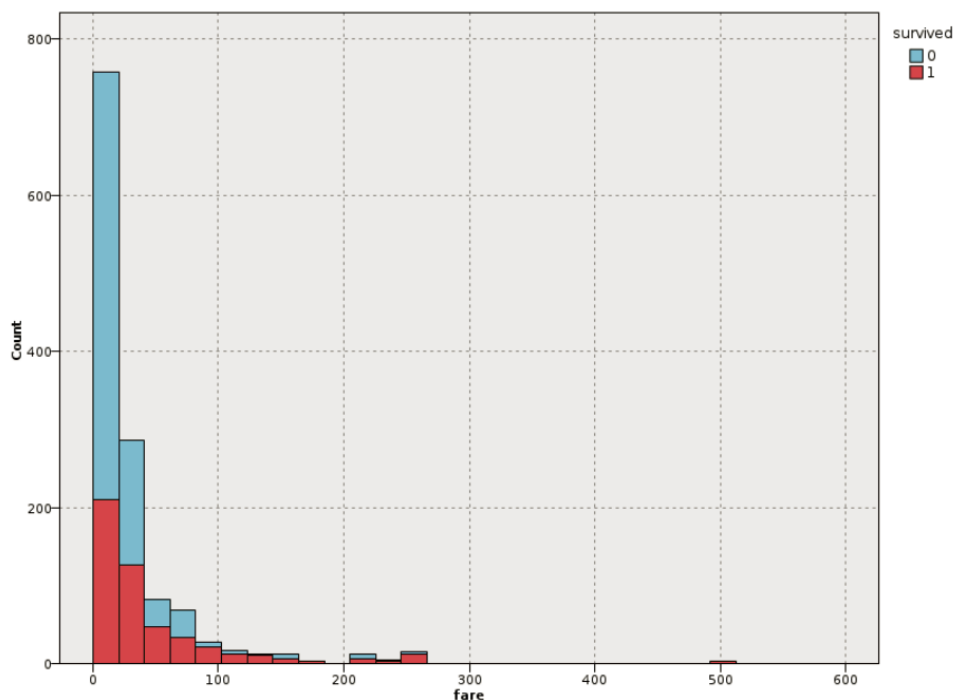
15. Right click on the **Histogram** node and select **Run**.



16. Click on the  icon next to the fare Histogram

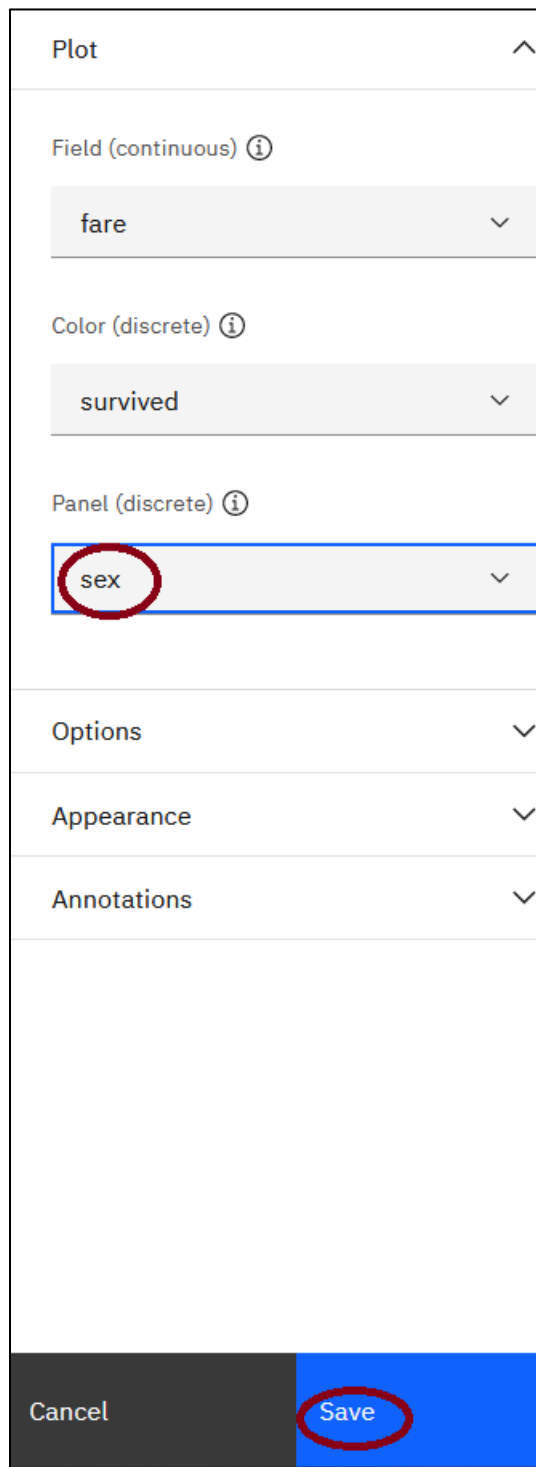


under the Outputs



17. We can see that the higher fares have a higher percentage of survival. We can also see that the histogram is skewed. Skewness will impact the effectiveness of some machine learning techniques. One way to deal with skewness is to do a logarithmic transformation of the data. We will do this transformation in the preparing the data for modeling section below.

18. You can view the above graph separately for male and female passengers. Return to the SPSS canvas by clicking on **Titanic SPSS** in the breadcrumb area. DoubleClick the Histogram icon. In the **Panel (discrete)** select sex, and the click **Save**.

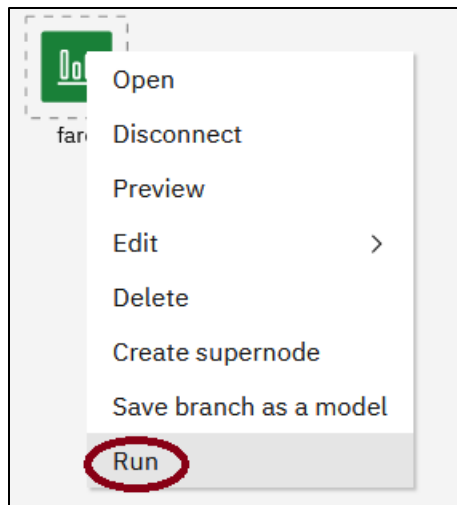



The image shows the 'Plot' dialog box in SPSS. It has a title bar 'Plot' with an expand/collapse arrow. The dialog is divided into several sections: 'Field (continuous)' with a dropdown menu showing 'fare'; 'Color (discrete)' with a dropdown menu showing 'survived'; 'Panel (discrete)' with a dropdown menu showing 'sex', which is highlighted with a red circle; 'Options' with a dropdown arrow; 'Appearance' with a dropdown arrow; and 'Annotations' with a dropdown arrow. At the bottom, there are two buttons: 'Cancel' on the left and 'Save' on the right, with the 'Save' button highlighted in blue and circled in red.

Section	Value
Field (continuous)	fare
Color (discrete)	survived
Panel (discrete)	sex
Options	
Appearance	
Annotations	

Buttons: Cancel, Save

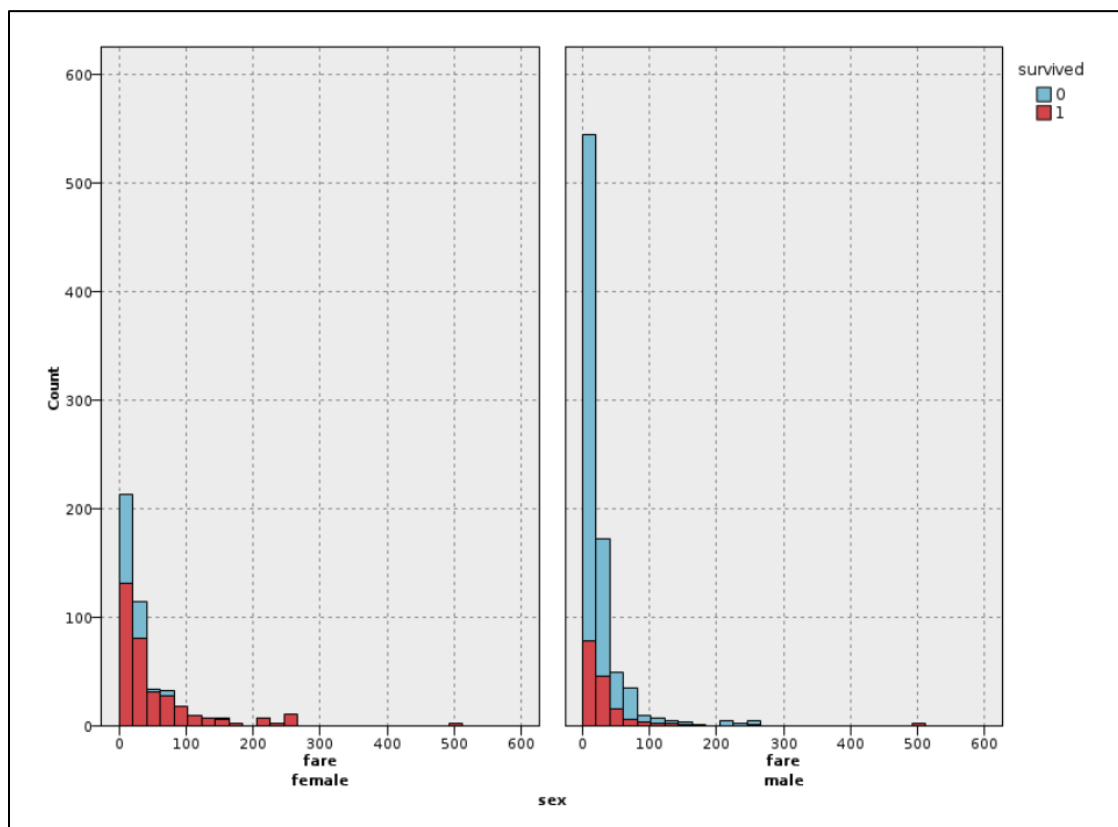
19. Right click on the Histogram and select **Run**.



20. Click on the  icon next to the fare Histogram under the Outputs tab at the right of the screen.



under the Outputs



21. Return to the SPSS canvas by clicking on **Titanic SPSS** in the breadcrumb area.


Step 2.4 Prepare the Data for Modeling

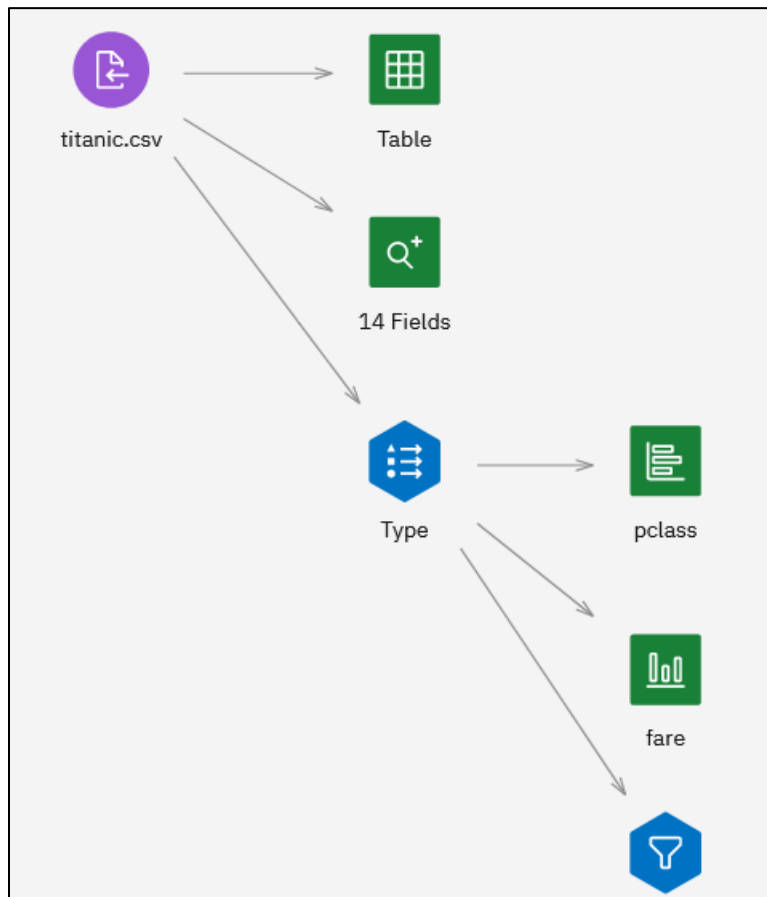
Based on our exploration of the data, there are several transformations that are needed to prepare the data for modeling. This section will introduce, the **Filter** node, the **Select** node, and the **Derive** node that will do the necessary transformations. The **Filter** and **Derive** nodes act on a field level, whereas the **Select** node acts on a record level.

Filter node – The **Filter** node performs two functions. It specifies fields that can be dropped. It also allows fields to be renamed. We will drop the fields cabin,boat,body, and home_dest.

Derive node – The **Derive** node modifies data values or creates new fields from one or more existing fields. We will use the derive node to do a logarithmic transformation of the fare field. We will also use this node to bin the age and fare fields.

Select node – The **Select** node is used to select or discard a subset of records from the data stream based on a specific condition. We will remove the rows where there is missing information in the fare, age, or embarked fields.

1. Add a **Filter** node to drop fields with many missing values. Add the **Filter** node by clicking on the **Field Operations** menu item in the Node palette and dragging the **Filter** node onto the canvas underneath the fare **Histogram** node. If the Node Palette is not visible, click on the Node Palette icon  first. Connect the **Type** node to the **Filter** node. The canvas should appear as below.



2. Double click on the **Filter** node. Click on the **Filter** dropdown. In the Filter panel, click on **Add Columns**.

Filter

Filter

Mode ⓘ

☒ Filter the selected fields

☐ Retain the selected fields (all other fields are filtered)

Filter Options ▾

Select Fields ⓘ

Remove Add Columns +

<input checked="" type="checkbox"/>	Field Name
-------------------------------------	------------

3. Click on the checkboxes adjacent to the **cabin**, **boat**, **body**, and **home_dest** fields, and then click on **OK**. Scroll down if necessary to locate these fields.

<input type="checkbox"/>	Field Name	Data Type
<input type="checkbox"/>	age	## double
<input type="checkbox"/>	sibsp	# integer
<input type="checkbox"/>	parch	# integer
<input type="checkbox"/>	ticket	abc string
<input type="checkbox"/>	fare	## double
<input checked="" type="checkbox"/>	cabin	abc string
<input type="checkbox"/>	embarked	abc string
<input checked="" type="checkbox"/>	boat	abc string
<input checked="" type="checkbox"/>	body	# integer
<input checked="" type="checkbox"/>	home_dest	abc string

Cancel


OK

4. Click **Save** on the Filter panel.

Filter

Mode ⓘ
☒ Filter the selected fields
☐ Retain the selected fields (all other fields are filtered)

Filter Options ▾

Select Fields ⓘ
Remove  Add Columns +

<input checked="" type="checkbox"/>	Field Name
<input checked="" type="checkbox"/>	cabin
<input checked="" type="checkbox"/>	boat
<input checked="" type="checkbox"/>	body
<input checked="" type="checkbox"/>	home_dest


Fields: 14 in, 4 filtered, 10 out

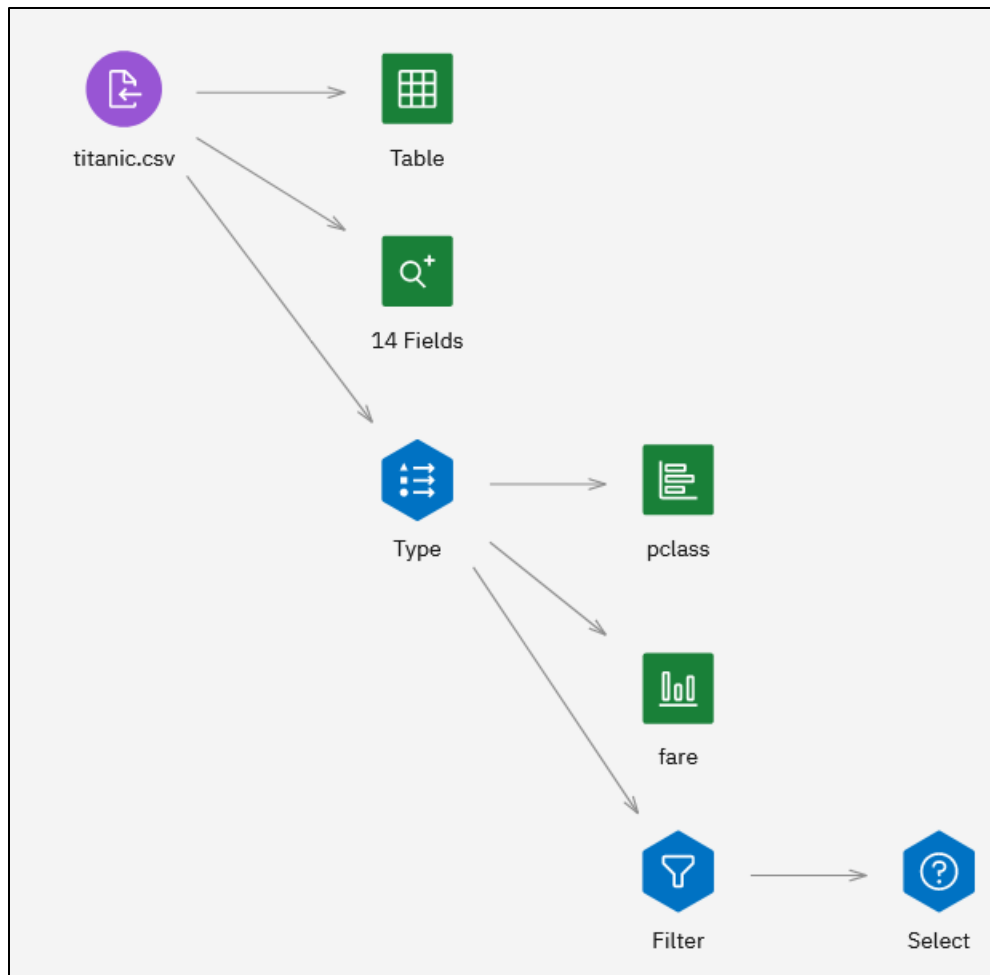
Rename ▾

Annotations ▾

Cancel

Save

5. Add a **Select** node by clicking on the **Record Operations** menu item in the Node palette, and then dragging the **Select** node to the canvas to the right of the **Filter** node. Connect the **Filter** node to the **Select** node. If the Node Palette is not visible, click on the Node Palette icon  first. The canvas should appear as below.



6. Double click on the **Select** node. Click on the **Settings** dropdown. In the **Select** panel, click on the **Discard** radio button, copy and paste (or type) the code shown below in the **Condition text box**, and then click **Save**.

@NULL (age) or embarked==" " or @NULL(fare)


Settings ^

Mode ⓘ

☐ Include

☒ Discard

Condition ⓘ




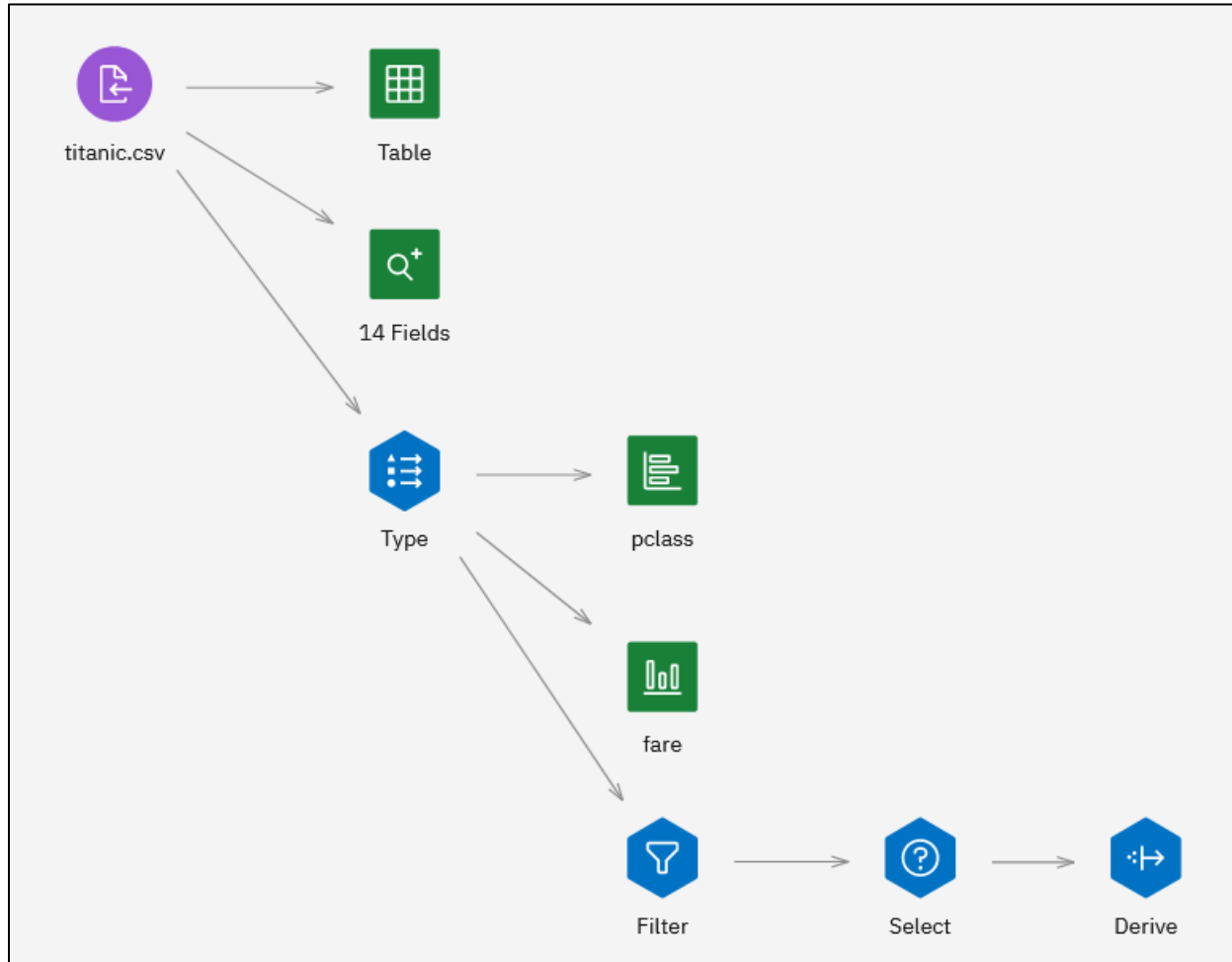
@NULL (age) or embarked==" or

< >

Annotations v

Cancel Save

7. Add a **Derive** node to the canvas by clicking on the **Field Operations** menu item in the Node palette, and then dragging the **Derive node** onto the canvas to the right of the **Select** node. If the Node Palette is not visible, click on the Node Palette icon  first. Connect the **Select** node to the **Derive** node. The canvas should appear as below.



8. Double click on the **Derive** node. Click on the **Settings** Dropdown. Click on the **Single** radio button, enter `log_fare` for the **Derive** field, select **Continuous** for the measurement, copy and paste (or type) the following code in the **Expression** text box, and click Save.

```
if (fare !=0) then log(fare)
```

```
else 0
```

```
endif
```

Settings

Mode ⓘ
☒ Single field
☐ Multiple fields

Derived Field Name ⓘ
log_fare

Derive As ⓘ
Formula

Measurement ⓘ
Continuous

Expression ⓘ


if (fare /=0) then log(fare)
else 0
endif

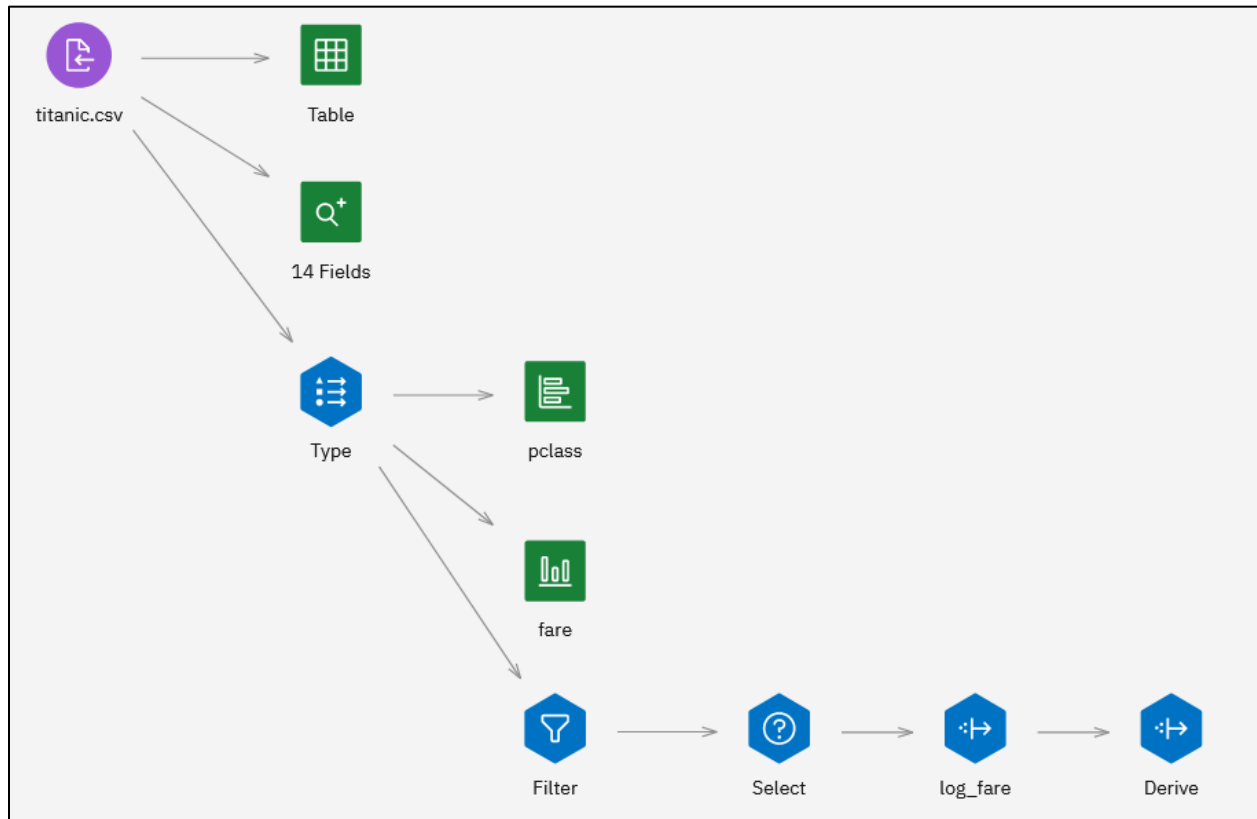
Annotations

Cancel

Save

9. Binning of continuous fields is a technique sometimes used in preparing data for modeling. We will bin the age field, and the log_fare field. Add a **Derive** node by clicking on the **Field Operations** menu item in the Node palette and dragging the **Derive** node on the canvas to the right of the log_fare **Derive** node.

If the Node Palette is not visible, click on the Node Palette icon  first. Connect the log_fare **Derive** node to the newly added **Derive** node. The canvas should appear as below.



10. Double click on the **Derive** node. Click on the **Settings** dropdown. Click on the **Single** radio button, enter age_bucket for the **Derive** field, select **Ordinal** for the **Measurement**, copy and paste the following code in the **Expression** text box, and then click **Save**.

```
if age >=0 and age < 6 then 0
else if age >=6 and age < 12 then 1
else if age>=12 and age< 18 then 2
else if age>=18 and age <40 then 3
else if age>=40 and age <65 then 4
else if age>=65 and age<80 then 5
else 6
endif
endif
endif
endif
endif
endif
```

Settings

Mode ⓘ
☒ Single field
☐ Multiple fields

Derived Field Name ⓘ
age_bucket

Derive As ⓘ
Formula

Measurement ⓘ
Ordinal


Expression ⓘ

```
if age >=0 and age < 6 the
else if age >=6 and age <
else if age>=12 and age< 1
else if age>=18 and age <4
else if age>=40 and age <6
else if age>=65 and age<80
```

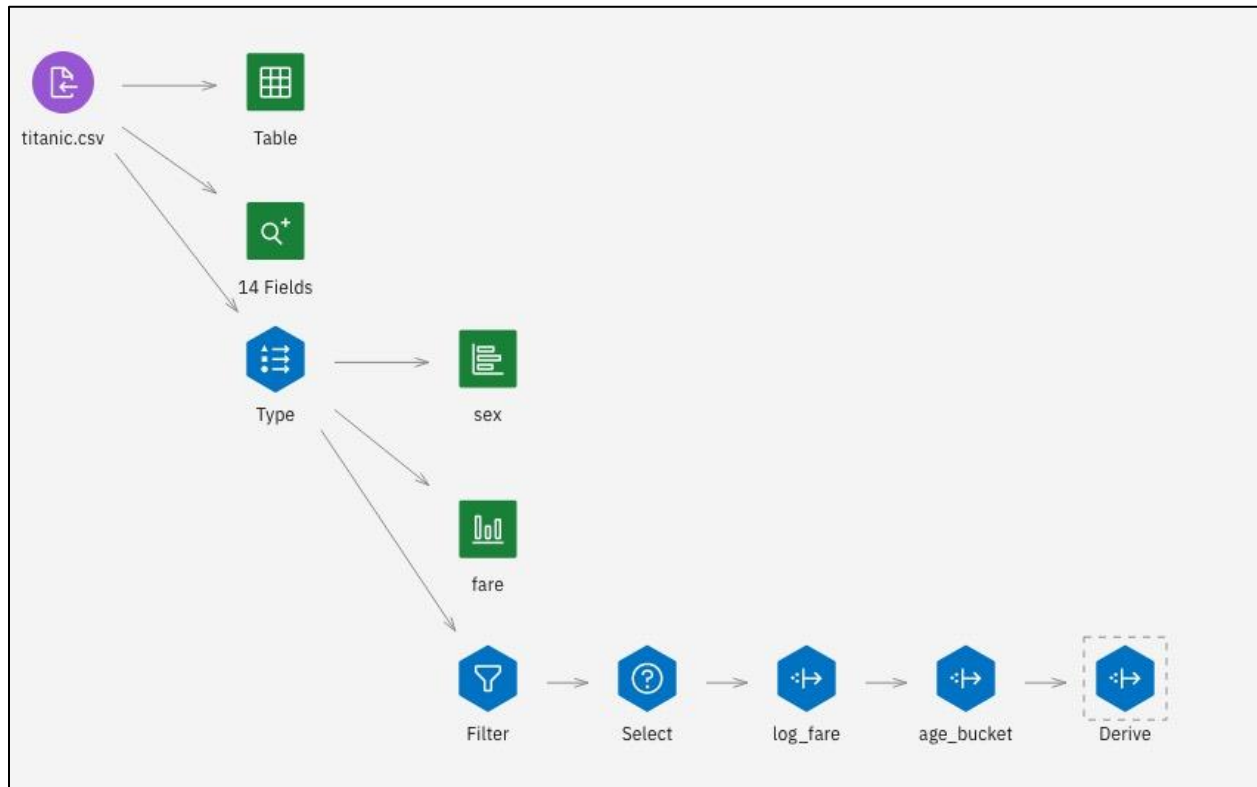
Annotations

Cancel

Save

11. Add a **Derive** node by clicking on the Field Operations menu item in the Node palette and dragging the **Derive** node onto the canvas to the right of the age_bucket **Derive** node. You can click on the **zoom to fit** icon  in the top right to fit the flow to the canvas.

Connect the age_bucket **Derive** node to the newly created **Derive** Node. The canvas should appear as below.



12. Double click the **Derive** node. In the **Derive** panel, click on the **Single** radio button, enter fare_bucket in the **Derive** field, click on Ordinal for the **Measurement**, copy and paste (or type) the following code in the **Expression** text box, and click on **Save**.

```
if log_fare < 0 then 0
else if log_fare > 8 then 9
else to_integer(log_fare)+1
endif
endif
```

Settings

Mode ⓘ
☒ Single field
☐ Multiple fields

Derived Field Name ⓘ
fare_bucket

Derive As ⓘ
Formula

Measurement ⓘ
Ordinal

Expression ⓘ

```
if log_fare < 0 then 0  
else if log_fare > 8 then 9  
else to_integer(log_fare)+1  
endif  
endif
```

Annotations

Cancel

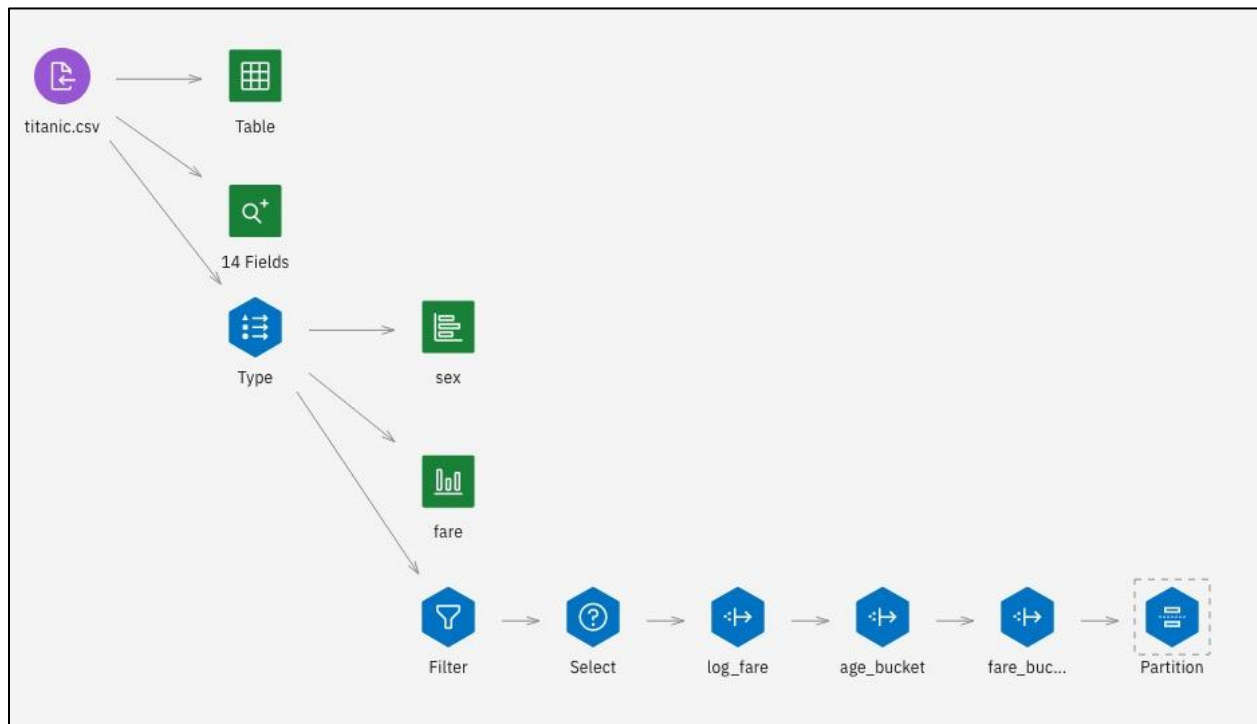
Save

Step 2.5 Modeling and Evaluation

Now that the data is prepared, we can start the modeling effort. First, we will add a **Partition** node to divide the data set into Training and Testing sets. In addition, a **Type** node is needed prior to modeling to type the new data fields that were created. Then we will add a **Logistic** node

and use the Training set to train the model. Finally, we will add an **Analysis** node to evaluate the results.

1. Add a **Partition** node by clicking on the Field Operations menu item in the Node palette and dragging the **Partition** node onto the canvas to the right of the fare_bucket **Derive** node. Connect the fare_bucket **Derive** node to the **Partition** node. The canvas should appear as below.



2. Double click on the Partition node. Set the **Training Partition** to 70 and the **Test Partition** to 30. Leave the other defaults and click on **Save**.

Settings

Derived Field Name ⓘ
Partition

Training Partition(%) ⓘ
70

Testing Partition(%) ⓘ
30

☐ Create validation partition

☒ Repeatable partition assignment

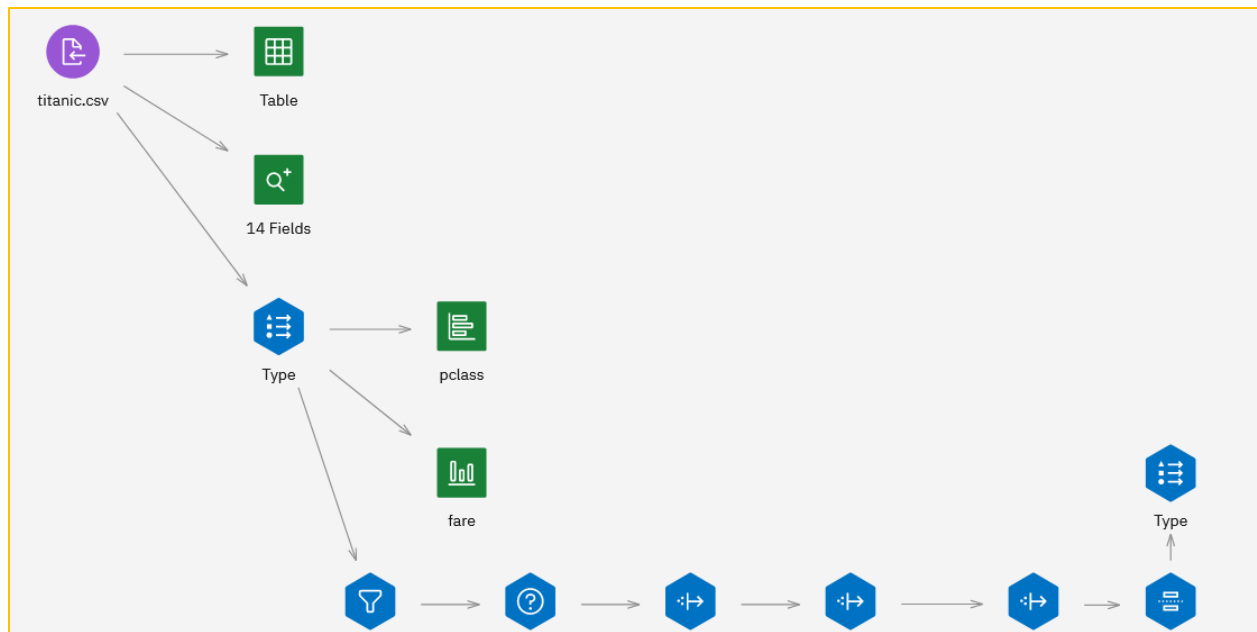
Seed [Generate](#) ⓘ
1234567

☐ Use unique field to assign partitions

Annotations

CancelSave

3. Add a **Type** node by clicking on the **Field Operations** in the Node palette and dragging the **Type** node onto the canvas above the **Partition** node. Connect the **Partition** node to the **Type** node. The canvas should appear as below.



4. Double click on the **Type** node. Click on **Read Values**.

Settings

Default Mode ⓘ
☒ Read metadata ☐ Pass (do not scan)

Type Operations

Read Values

Clear All Values

Find in column Field

<input type="checkbox"/>	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# pclass	Ordinal	Input	Pass	1, 2, 3	None
<input type="checkbox"/>	# survived	Flag	Input	Pass	0, 1	None
<input type="checkbox"/>	abc name	Typeless	None	Pass		None
<input type="checkbox"/>	abc sex	Flag	Input	Pass	female, male	None
<input type="checkbox"/>	# age	Continuous	Input	Pass	0.1667, 80.0	None
<input type="checkbox"/>	# sibsp	Continuous	Input	Pass	0, 8	None
<input type="checkbox"/>	# parch	Continuous	Input	Pass	0, 9	None
<input type="checkbox"/>	abc ticket	Typeless	None	Pass		None

Format

Annotations

Cancel

Save

- For the log_fare, select **Continuous** for the **Measurement**. For the fare_bucket field, select **Ordinal** for the **Measurement**, and for the age_bucket, select **Ordinal** for the **Measurement**, (note these values should already be set correctly).

Settings

Default Mode ⓘ
☒ Read metadata ☐ Pass (do not scan)

Type Operations

Read Values Clear All Values

Find in column Field

<input type="checkbox"/>	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# parch	Continuous ▾	Input ▾	Instantiated	0, 9	None ▾ ⚙
<input type="checkbox"/>	abc ticket	Typeless ▾	None ▾	Pass ▾		None ▾ ⚙
<input type="checkbox"/>	# fare	Continuous ▾	Input ▾	Instantiated	0.0, 512.3292	None ▾ ⚙
<input type="checkbox"/>	abc embarke	Nominal ▾	Input ▾	Instantiated	C, Q, S	None ▾ ⚙
<input type="checkbox"/>	# log_fare	Continuous ▾	Input ▾	Instantiated	0.0, 6.23896738...	None ▾ ⚙
<input type="checkbox"/>	# age_buc	Ordinal ▾	Input ▾	Instantiated	0, 1, 2, 3, 4, 5, 6	None ▾ ⚙
<input type="checkbox"/>	# fare_buc	Ordinal ▾	Input ▾	Instantiated	1, 2, 3, 4, 5, 6, 7	None ▾ ⚙
<input type="checkbox"/>	abc Partition	Nominal ▾	Partition ▾	Instantiated	1_Training, 2_Te...	None ▾ ⚙

Format

Annotations

Cancel Save

- Update the **Role** of the following **Fields**: survived→ Target, age→None, fare→None, log_fare→None and click **Save**.

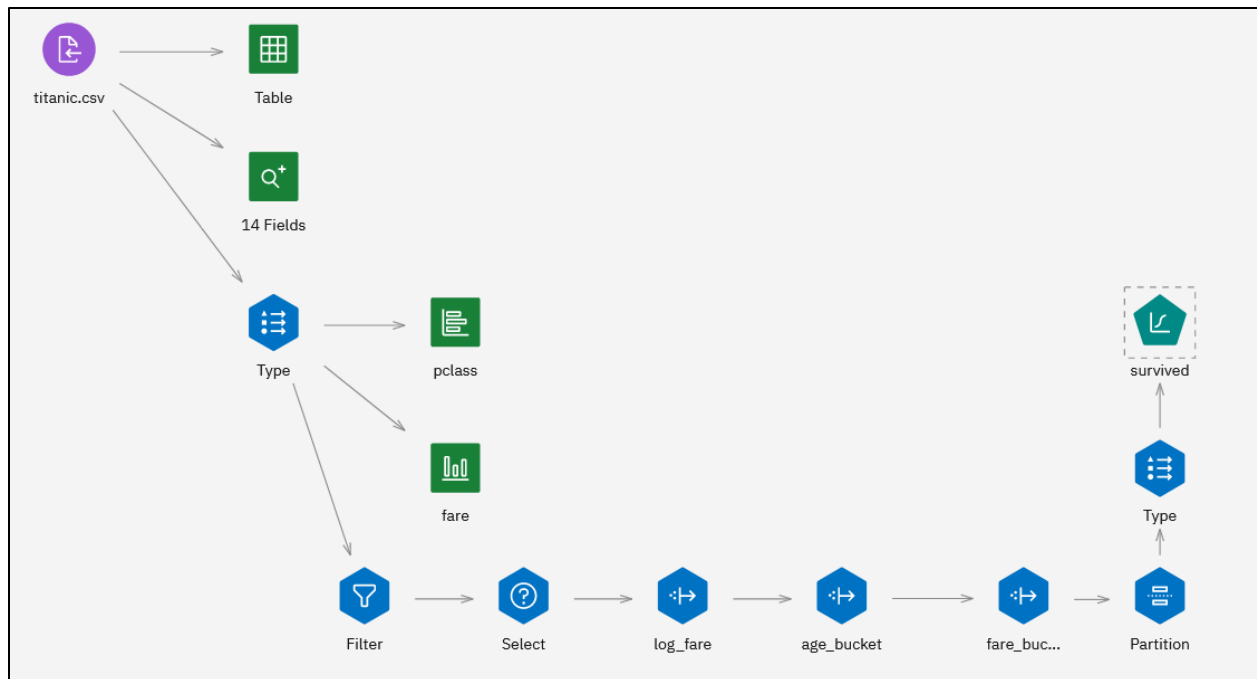
<input type="checkbox"/>	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# pclass	Ordinal	Input	Instantiated	1, 2, 3	None
<input type="checkbox"/>	# survived	Flag	Target	Instantiated	0, 1	None
<input type="checkbox"/>	abc name	Typeless	None	Pass		None
<input type="checkbox"/>	abc sex	Flag	Input	Instantiated	female, male	None
<input type="checkbox"/>	# age	Continuous	None	Instantiated	0.1667, 80.0	None
<input type="checkbox"/>	# sibsp	Continuous	Input	Instantiated	0, 8	None
<input type="checkbox"/>	# parch	Continuous	Input	Instantiated	0, 9	None
<input type="checkbox"/>	abc ticket	Typeless	None	Pass		None

<input type="checkbox"/>	Field	Measure	Role	Value Mode	Values	Check
<input type="checkbox"/>	# parch	Continuous	Input	Instantiated	0, 9	None
<input type="checkbox"/>	abc ticket	Typeless	None	Pass		None
<input type="checkbox"/>	# fare	Continuous	None	Instantiated	0.0, 512.3292	None
<input type="checkbox"/>	abc embarke	Nominal	Input	Instantiated	C, Q, S	None
<input type="checkbox"/>	# log_fare	Continuous	None	Instantiated	0.0, 6.23896738...	None
<input type="checkbox"/>	# age_bucl	Ordinal	Input	Instantiated	0, 1, 2, 3, 4, 5, 6	None
<input type="checkbox"/>	# fare_buc	Ordinal	Input	Instantiated	1, 2, 3, 4, 5, 6, 7	None
<input type="checkbox"/>	abc Partition	Nominal	Partition	Instantiated	1_Training, 2_Te...	None

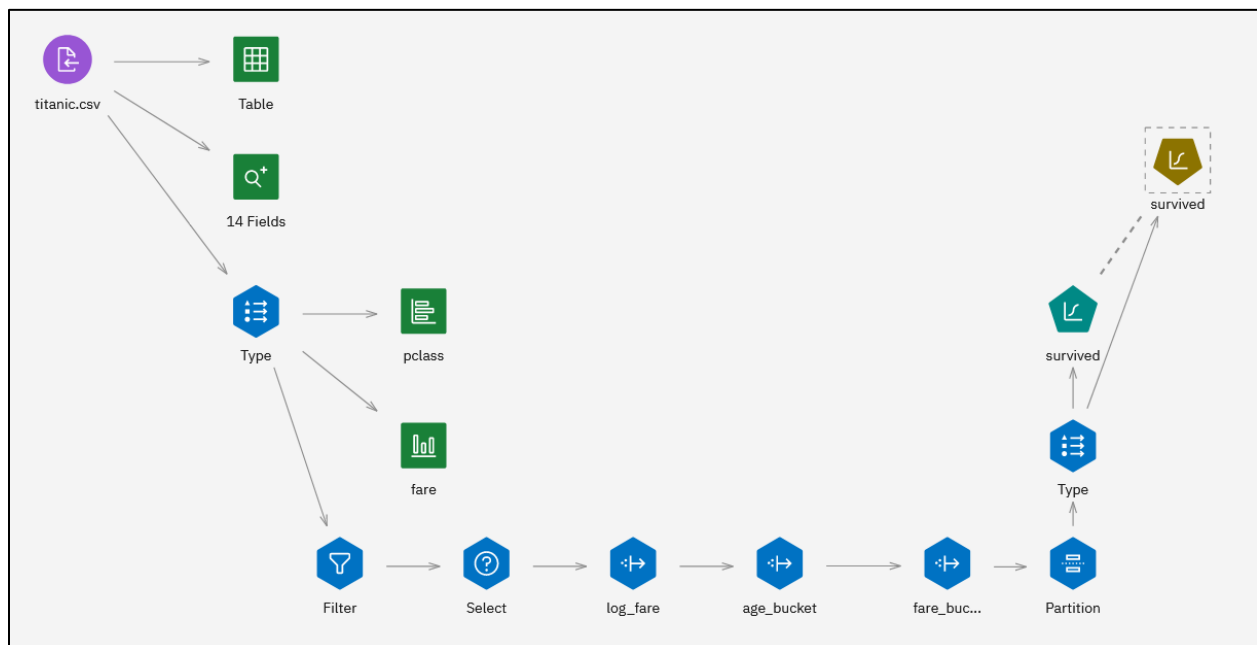
Format
Annotations

Cancel
Save

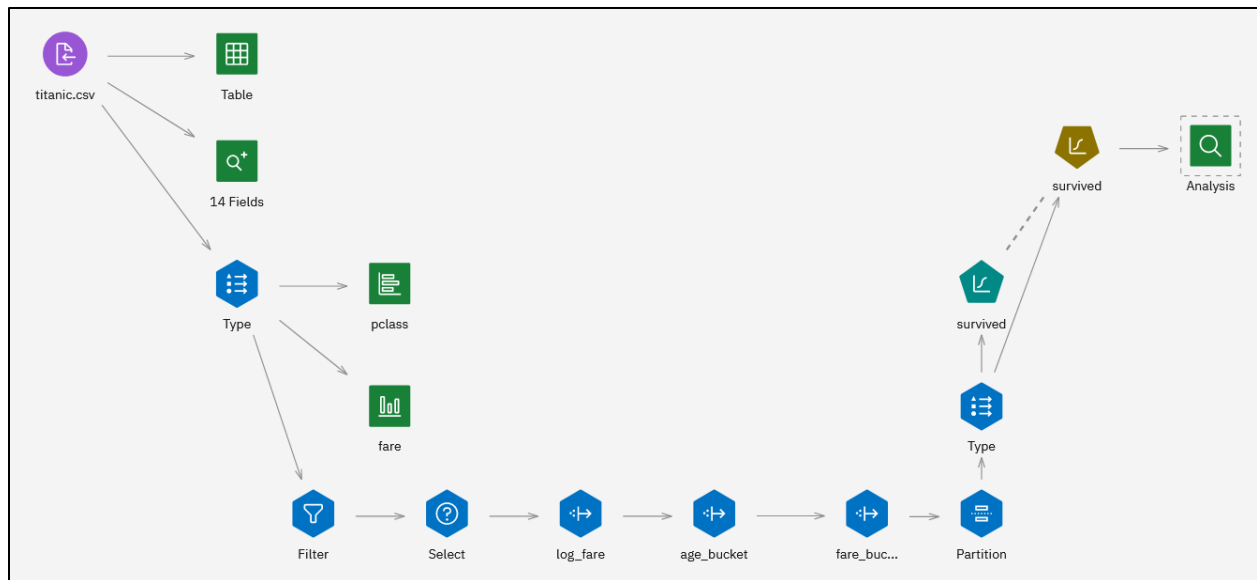
7. Add a **Logistic** node by clicking on the **Modeling** menu item in the Node palette and dragging the **Logistic** node onto the canvas above the **Type** node. Connect the **Type** node to the **Logistic** node. The canvas should appear as below.



8. Right click on the **Logistic** node and then click **Run**. A **Logistic** “nugget” will be created” connected by a dotted line to the **Logistic** node. Note, it may be hidden under another node. Drag the nugget and place it above the **Logistic** node. The canvas should appear as below.



9. Add an **Analysis** node by clicking on the **Outputs** menu item in the Node palette and dragging the **Analysis** node onto the canvas above the nugget icon. Connect the nugget icon to the **Analysis** node. The canvas should appear as below.



10. Double click on the Analysis node. Click on the **Settings** dropdown. Click on the **Evaluation metric** checkbox and click on **Save**.

Settings

^

☐ Coincidence matrices (for symbolic targets)

☐ Performance evaluation

☒ Evaluation metric (AUC & Gini, binary classifiers only)

☐ Confidence figures (if available)

Threshold for pct. correct ⓘ

90

Improve accuracy multiplier ⓘ

2

Find predicted/predictor fields using ⓘ

☒ Model output field metadata

☐ Field name format (for example, '\$<x>-<target field>')


☒ Separate by partition

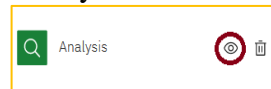
☐ User defined analysis

Configure Analysis +

Cancel

Save

11. Right click on the Analysis node and select Run. After completion, click on the  icon



next to **Analysis** in the Outputs tab on the right side of the screen. The results should be similar to those shown below.

[← Return to flow](#)

View Output: Analysis

Results for output field survived

Individual Models

Comparing \$L-survived with survived

'Partition'	1_Training	2_Testing
Correct	719 79.01%	321 80.45%
Wrong	191 20.99%	78 19.55%
Total	910	399

Evaluation Metrics

'Partition'	1_Training	2_Testing
Model	AUC Gini	AUC Gini
\$L-survived	0.841 0.682	0.872 0.743

12. Click on **Return to flow**.

[← Return to flow](#)

View Output: Analysis

Results for output field survived

Individual Models

Comparing \$L-survived with survived

'Partition'	1_Training	2_Testing
Correct	719 79.01%	321 80.45%
Wrong	191 20.99%	78 19.55%
Total	910	399

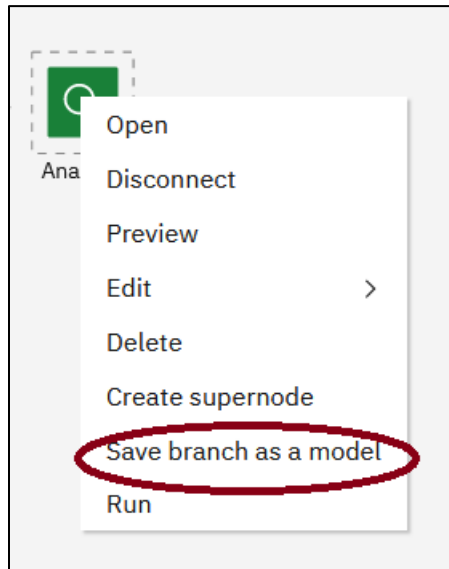
Evaluation Metrics

'Partition'	1_Training	2_Testing
Model	AUC Gini	AUC Gini
\$L-survived	0.841 0.682	0.872 0.743

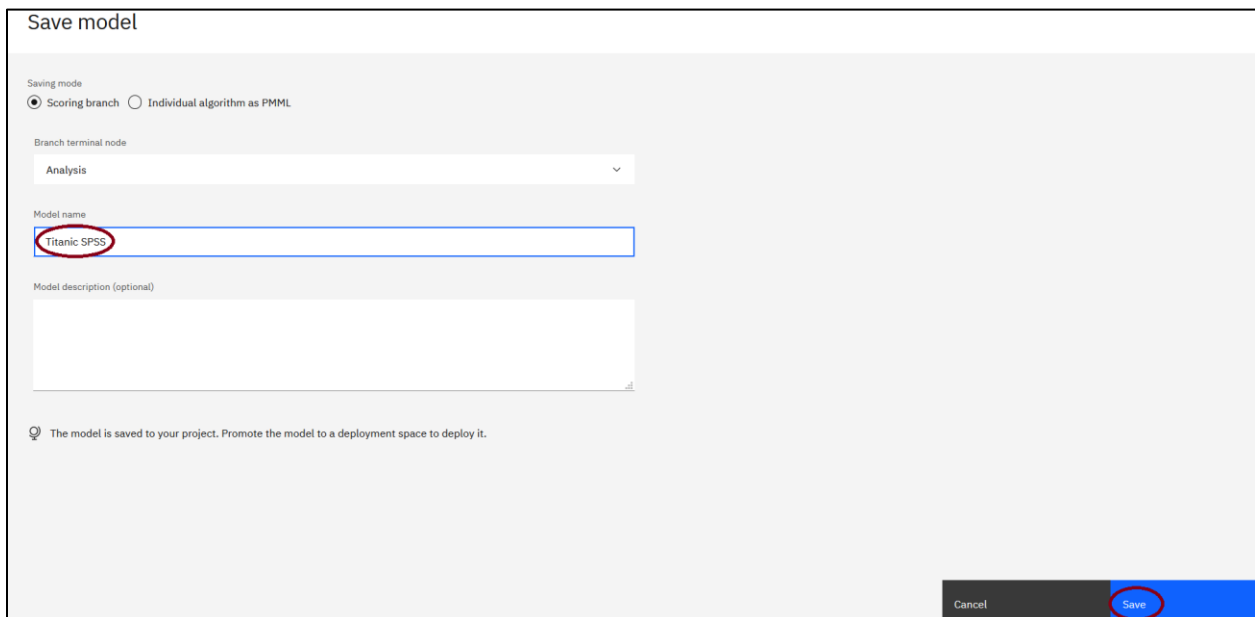
Step 2.6 Saving a Model

Now that we have created and evaluated a model, we will save the model as an asset. This saved model can be deployed at a future date, removing the need to recreate the same model from scratch.

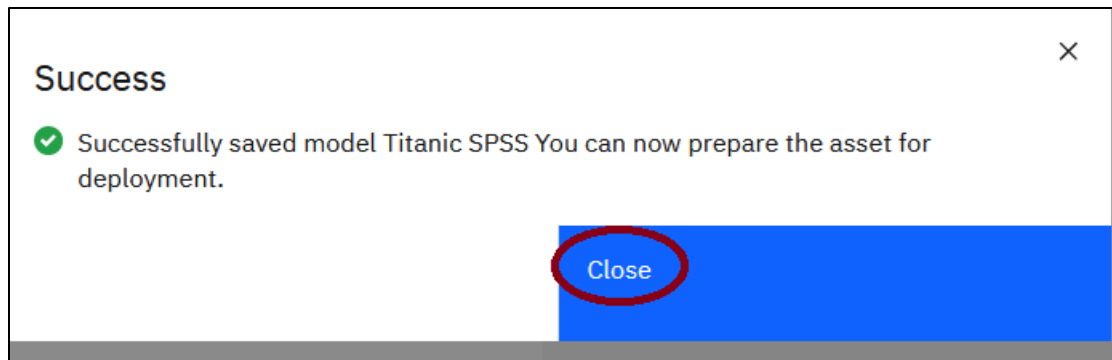
1. Right click on the Analysis node and then click on **Save branch as a model**.



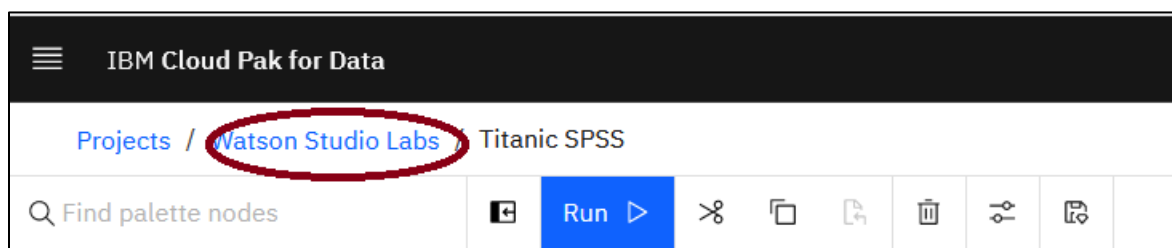
2. Type in “**Titanic SPSS**” as the Model Name and click **Save**.

A screenshot of a 'Save model' dialog box. The title bar says 'Save model'. Inside, there's a 'Saving mode' section with two radio buttons: 'Scoring branch' (selected) and 'Individual algorithm as PMML'. Below that is a 'Branch terminal node' dropdown menu showing 'Analysis'. The 'Model name' field contains the text 'Titanic SPSS' and is circled in red. Below that is a 'Model description (optional)' text area. At the bottom right, there are two buttons: 'Cancel' and 'Save' (which is circled in red). A small icon and text at the bottom left state: 'The model is saved to your project. Promote the model to a deployment space to deploy it.'

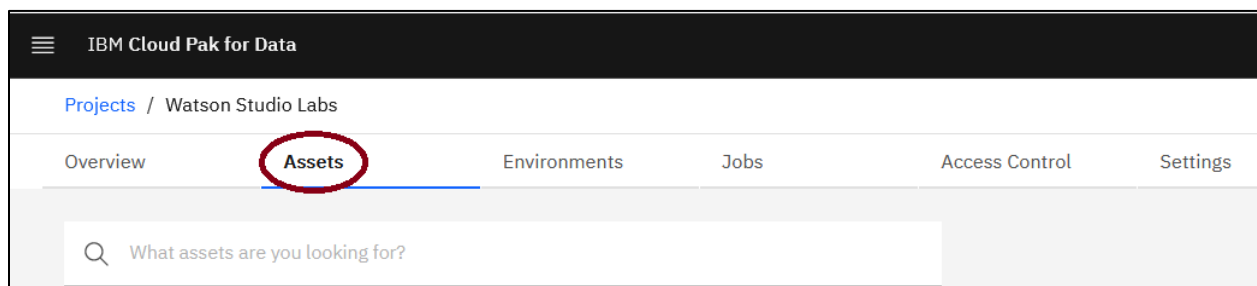
3. Click **Close**.



4. Click on Watson Studio Labs to return to the Project.



5. Click on the Assets tab if necessary.



6. Note that the model you built is now saved as an asset and the work you have completed can be easily reused in the future.

Models			
Watson Machine Learning models			
			New model from file +
Name	Type	Software specification	Last modified ↓
Titanic SPSS	spss-modeler_18.2	spss-modeler_18.2	Oct 26, 2020