

# WM\_Sales\_Forecast



**Version**

1.0

**Author**

jayasurya000008

**Date**

2023-12-01 16:49:37

## Table of contents

<b>I.</b>	<b>Project general description</b>	<b>4</b>
<b>II.</b>	<b>Graphical Representation of the Flow</b>	<b>5</b>
<b>III.</b>	<b>Datasets</b>	<b>6</b>
1.	{{ \$dataset.name }}	Error! Bookmark not defined.
a)	Status of the dataset {{ \$dataset.name }}	Error! Bookmark not defined.
b)	Related Recipes	20
c)	Schema	21
<b>IV.</b>	<b>Managed folders</b>	Error! Bookmark not defined.
1.	{{ \$folder.name }}	Error! Bookmark not defined.
a)	Status	Error! Bookmark not defined.
b)	Derived datasets	Error! Bookmark not defined.
c)	Related Recipes	Error! Bookmark not defined.
<b>V.</b>	<b>Recipes</b>	<b>22</b>
1.	{{ \$recipe.name }}	Error! Bookmark not defined.
a)	Status	29
b)	Inputs / Outputs	29
c)	Prepare Steps	24
d)	Join	25
e)	Data Relationships	Error! Bookmark not defined.
f)	Feature Transformations	Error! Bookmark not defined.
g)	Code	Error! Bookmark not defined.
h)	Windows	Error! Bookmark not defined.
i)	Sorted Columns	Error! Bookmark not defined.
j)	Top rows and sorting direction	Error! Bookmark not defined.
k)	Parameters	Error! Bookmark not defined.
<b>VI.</b>	<b>Saved Models</b>	<b>30</b>
1.	{{ \$model.name }}	Error! Bookmark not defined.
a)	Status	31
b)	Related Recipes	31
<b>VII.</b>	<b>Model Evaluation Stores</b>	Error! Bookmark not defined.
1.	{{ \$mes.name }}	Error! Bookmark not defined.
a)	Related Recipes	Error! Bookmark not defined.
<b>VIII.</b>	<b>Labeling Tasks</b>	Error! Bookmark not defined.
1.	{{ \$labeling_task.name }}	Error! Bookmark not defined.
a)	Status	Error! Bookmark not defined.
b)	Inputs / Outputs	Error! Bookmark not defined.
<b>IX.</b>	<b>Deployment and Monitoring</b>	<b>32</b>

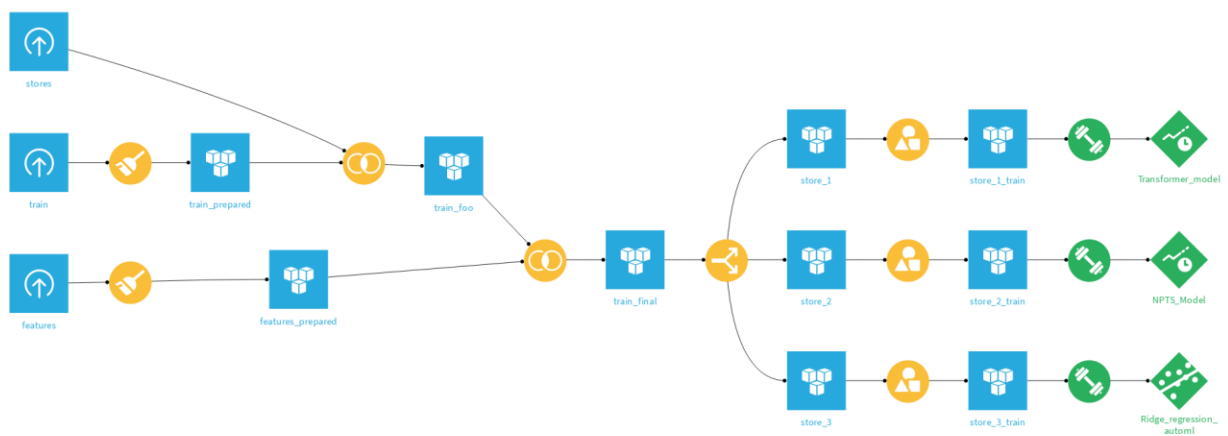
A.	Implementation Details	32
B.	Version Control	32

## I. Project general description

Project references for the created flow:

- Project name: **WM\_Sales\_Forecast**
- Project key: **WM\_SALES\_FORECAST**
- Project short description:  
*The project \*WM\_Sales\_Forecast\* was created by jayasurya00008 on Nov 27th 2023*
- Status: **Sandbox**
- Current git branch: **master**
- Created by **jayasurya00008** on the 2023-11-27 20:03:55
- Last Modification by **jayasurya00008** on the 2023-12-01 16:47:28

## II. Graphical Representation of the Flow



### III. Datasets

List all the datasets in the flow.

#### 1. stores

Summarized information about your dataset stores:

##### a) Status of the dataset stores

Characteristics of the dataset stores:

- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **45**
- Created by **jayasurya00008** on the 2023-11-27 20:07:39
- Last Modification by **jayasurya00008** on the 2023-11-30 19:38:59

##### b) Related Recipes

The dataset stores relates to the following recipes.

##### Successor Recipes

List of recipes with stores as an input:

- compute\_train\_foo ([Join](#))

##### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Type	string	

Size	string	
------	--------	--

## 2. train

Summarized information about your dataset train:

### a) Status of the dataset train

Characteristics of the dataset train:

- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-27 20:07:37
- Last Modification by **jayasurya00008** on the 2023-11-27 20:07:38

### b) Related Recipes

The dataset train relates to the following recipes.

#### Successor Recipes

List of recipes with train as an input:

- compute\_train\_prepared ([Prepare](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	string	
Dept	string	
Date	string	
Weekly_Sales	string	
IsHoliday	string	

### 3. train\_prepared

Summarized information about your dataset train\_prepared:

#### a) Status of the dataset train\_prepared

Characteristics of the dataset train\_prepared:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **2.4 MB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-27 20:49:52
- Last Modification by **jayasurya00008** on the 2023-11-30 19:39:11

#### b) Related Recipes

The dataset train\_prepared relates to the following recipes.

##### Parent Recipe

List of recipes with train\_prepared as an output:

- compute\_train\_prepared ([Prepare](#))

##### Successor Recipes

List of recipes with train\_prepared as an input:

- compute\_train\_foo ([Join](#))

#### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Dept	bigint	
Date	string	
Weekly_Sales	double	



IsHoliday	boolean	
-----------	---------	--

#### 4. train\_foo

Summarized information about your dataset train\_foo:

##### a) Status of the dataset train\_foo

Characteristics of the dataset train\_foo:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **2.4 MB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 19:11:22
- Last Modification by **jayasurya00008** on the 2023-11-30 19:41:07

##### b) Related Recipes

The dataset train\_foo relates to the following recipes.

##### Parent Recipe

List of recipes with train\_foo as an output:

- compute\_train\_foo ([Join](#))

##### Successor Recipes

List of recipes with train\_foo as an input:

- compute\_train\_final ([Join](#))

##### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Type	string	

<b>Size</b>	string	
<b>Dept</b>	bigint	
<b>Date</b>	date	
<b>Weekly_Sales</b>	double	
<b>IsHoliday</b>	boolean	

## 5. features

Summarized information about your dataset features:

### a) Status of the dataset features

Characteristics of the dataset features:

- Type: **Uploaded Files**
- Connection: **dataiku-managed-storage**
- Total Size: **Not computed**
- Record Count: **8,190**
- Created by **jayasurya00008** on the 2023-11-30 19:22:07
- Last Modification by **jayasurya00008** on the 2023-11-30 19:22:08

### b) Related Recipes

The dataset features relates to the following recipes.

#### Successor Recipes

List of recipes with features as an input:

- compute\_features\_prepared (*Prepare*)

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<b>Store</b>	string	
<b>Date</b>	string	

Temperature	string	
Fuel_Price	string	
MarkDown1	string	
MarkDown2	string	
MarkDown3	string	
MarkDown4	string	
MarkDown5	string	
CPI	string	
Unemployment	string	
IsHoliday	string	

## 6. features\_prepared

Summarized information about your dataset features\_prepared:

### a) Status of the dataset features\_prepared

Characteristics of the dataset features\_prepared:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **73.8 KB**
- Record Count: **8,190**
- Created by **jayasurya00008** on the 2023-11-30 19:22:54
- Last Modification by **jayasurya00008** on the 2023-11-30 19:29:02

### b) Related Recipes

The dataset features\_prepared relates to the following recipes.

#### Parent Recipe

List of recipes with features\_prepared as an output:

- compute\_features\_prepared ([Prepare](#))

## Successor Recipes

List of recipes with features\_prepared as an input:

- compute\_train\_final ([Join](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	bigint	
Date	date	
Temperature	double	
Fuel_Price	double	
CPI	string	
Unemployment	string	
IsHoliday	boolean	

## 7. train\_final

Summarized information about your dataset train\_final:

### a) Status of the dataset train\_final

Characteristics of the dataset train\_final:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **2.5 MB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 19:33:16
- Last Modification by **jayasurya00008** on the 2023-11-30 20:18:56

### b) Related Recipes

The dataset train\_final relates to the following recipes.

## Parent Recipe

List of recipes with train\_final as an output:

- compute\_train\_final ([Join](#))

## Successor Recipes

List of recipes with train\_final as an input:

- split\_train\_final ([Split](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Type	string	
Size	int	
Dept	int	
Date	date	
Weekly_Sales	double	
IsHoliday	boolean	
Temperature	double	
Fuel_Price	double	
CPI	double	
Unemployment	double	

## 8. store\_2

Summarized information about your dataset store\_2:

### a) Status of the dataset store\_2

Characteristics of the dataset store\_2:

- Type: **Amazon S3**

- Connection: **dataiku-managed-storage**
- Total Size: **67.1 KB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 20:35:33
- Last Modification by **jayasurya00008** on the 2023-11-30 20:36:15

## b) Related Recipes

The dataset store\_2 relates to the following recipes.

### Parent Recipe

List of recipes with store\_2 as an output:

- split\_train\_final (*Split*)

### Successor Recipes

List of recipes with store\_2 as an input:

- compute\_store\_2\_train (*Group*)

## c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
<b>Store</b>	int	
<b>Type</b>	string	
<b>Size</b>	int	
<b>Dept</b>	int	
<b>Date</b>	date	
<b>Weekly_Sales</b>	double	
<b>IsHoliday</b>	boolean	
<b>Temperature</b>	double	
<b>Fuel_Price</b>	double	
<b>CPI</b>	double	

Unemployment	double	
--------------	--------	--

## 9. store\_2\_train

Summarized information about your dataset store\_2\_train:

### a) Status of the dataset store\_2\_train

Characteristics of the dataset store\_2\_train:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **7.4 KB**
- Record Count: **143**
- Created by **jayasurya00008** on the 2023-11-30 20:46:09
- Last Modification by **jayasurya00008** on the 2023-11-30 20:46:09

### b) Related Recipes

The dataset store\_2\_train relates to the following recipes.

#### Parent Recipe

List of recipes with store\_2\_train as an output:

- compute\_store\_2\_train ([Group](#))

#### Successor Recipes

List of recipes with store\_2\_train as an input:

- train\_NPTS\_Model ([Prediction Training](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Date	date	
Store_distinct	bigint	

Type_first	string	
Size_avg	double	
Dept_first	int	
Weekly_Sales_avg	double	
IsHoliday_first	boolean	
Temperature_avg	double	
Fuel_Price_avg	double	
CPI_avg	double	
Unemployment_avg	double	
count	bigint	

## 10. store\_3

Summarized information about your dataset store\_3:

### a) Status of the dataset store\_3

Characteristics of the dataset store\_3:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **57.4 KB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 20:35:40
- Last Modification by **jayasurya00008** on the 2023-11-30 20:36:15

### b) Related Recipes

The dataset store\_3 relates to the following recipes.

#### Parent Recipe

List of recipes with store\_3 as an output:

- split\_train\_final ([Split](#))



## Successor Recipes

List of recipes with store\_3 as an input:

- compute\_store\_3\_train (*Group*)

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Type	string	
Size	int	
Dept	int	
Date	date	
Weekly_Sales	double	
IsHoliday	boolean	
Temperature	double	
Fuel_Price	double	
CPI	double	
Unemployment	double	

## 11. store\_3\_train

Summarized information about your dataset store\_3\_train:

### a) Status of the dataset store\_3\_train

Characteristics of the dataset store\_3\_train:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **7.3 KB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 20:46:49

- Last Modification by **jayasurya00008** on the 2023-11-30 20:46:49

## b) Related Recipes

The dataset store\_3\_train relates to the following recipes.

### Parent Recipe

List of recipes with store\_3\_train as an output:

- compute\_store\_3\_train (*Group*)

### Successor Recipes

List of recipes with store\_3\_train as an input:

- train\_Ridge\_regression\_automl (*Prediction Training*)

## c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Date	date	
Store_distinct	bigint	
Type_first	string	
Size_avg	double	
Dept_first	int	
Weekly_Sales_avg	double	
IsHoliday_first	boolean	
Temperature_avg	double	
Fuel_Price_avg	double	
CPI_avg	double	
Unemployment_avg	double	
count	bigint	

## 12. store\_1

Summarized information about your dataset store\_1:

### a) Status of the dataset store\_1

Characteristics of the dataset store\_1:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **66.7 KB**
- Record Count: **Not computed**
- Created by **jayasurya00008** on the 2023-11-30 20:35:26
- Last Modification by **jayasurya00008** on the 2023-11-30 20:36:15

### b) Related Recipes

The dataset store\_1 relates to the following recipes.

#### Parent Recipe

List of recipes with store\_1 as an output:

- split\_train\_final ([Split](#))

#### Successor Recipes

List of recipes with store\_1 as an input:

- compute\_store\_1\_train ([Group](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Store	int	
Type	string	
Size	int	
Dept	int	

<b>Date</b>	date	
<b>Weekly_Sales</b>	double	
<b>IsHoliday</b>	boolean	
<b>Temperature</b>	double	
<b>Fuel_Price</b>	double	
<b>CPI</b>	double	
<b>Unemployment</b>	double	

### 13. store\_1\_train

Summarized information about your dataset store\_1\_train:

#### a) Status of the dataset store\_1\_train

Characteristics of the dataset store\_1\_train:

- Type: **Amazon S3**
- Connection: **dataiku-managed-storage**
- Total Size: **7.4 KB**
- Record Count: **143**
- Created by **jayasurya00008** on the 2023-11-30 20:41:27
- Last Modification by **jayasurya00008** on the 2023-11-30 20:44:39

#### b) Related Recipes

The dataset store\_1\_train relates to the following recipes.

##### Parent Recipe

List of recipes with store\_1\_train as an output:

- compute\_store\_1\_train ([Group](#))

##### Successor Recipes

List of recipes with store\_1\_train as an input:

- train\_Transformer\_model ([Prediction Training](#))

### c) Schema

Schema details of the dataset.

Column Name	Column Type	Description
Date	date	
Store_distinct	bigint	
Type_first	string	
Size_avg	double	
Dept_first	int	
Weekly_Sales_avg	double	
IsHoliday_first	boolean	
Temperature_avg	double	
Fuel_Price_avg	double	
CPI_avg	double	
Unemployment_avg	double	
count	bigint	

## IV. Recipes

List all the recipes in the flow.

### 1. compute\_train\_prepared

Summarized information about the recipe compute\_train\_prepared.

#### a) Status

Characteristics of the recipe compute\_train\_prepared:

- Type: **Prepare**
- Created by **jayasurya00008** on the 2023-11-27 20:49:52
- Last Modification by **jayasurya00008** on the 2023-11-30 19:35:05

#### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset train ([Uploaded Files](#))

List of the outputs of the recipe:

- Dataset train\_prepared ([Amazon S3](#))

#### c) Prepare Steps

Type	Comment
MeasureNormalize	
DateParser	

### 2. compute\_train\_foo

Summarized information about the recipe compute\_train\_foo.

### a) Status

Characteristics of the recipe compute\_train\_foo:

- Type: **Join**
- Created by **jayasurya00008** on the 2023-11-30 19:11:22
- Last Modification by **jayasurya00008** on the 2023-11-30 19:11:59

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset stores [\(Uploaded Files\)](#)
- Dataset train\_prepared [\(Amazon S3\)](#)

List of the outputs of the recipe:

- Dataset train\_foo [\(Amazon S3\)](#)

### c) Join

Inner stores - train\_prepared And

Left input	Condition	Right input
Store	=	Store

## 3. compute\_features\_prepared

Summarized information about the recipe compute\_features\_prepared.

### a) Status

Characteristics of the recipe compute\_features\_prepared:

- Type: **Prepare**
- Created by **jayasurya00008** on the 2023-11-30 19:22:54
- Last Modification by **jayasurya00008** on the 2023-11-30 19:32:20

## b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset features ([Uploaded Files](#))

List of the outputs of the recipe:

- Dataset features\_prepared ([Amazon S3](#))

## c) Prepare Steps

Type	Comment
ColumnsSelector	
FindReplace	
FindReplace	

## 4. compute\_train\_final

Summarized information about the recipe compute\_train\_final.

### a) Status

Characteristics of the recipe compute\_train\_final:

- Type: **Join**
- Created by **jayasurya00008** on the 2023-11-30 19:33:16
- Last Modification by **jayasurya00008** on the 2023-11-30 19:36:21

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset train\_foo ([Amazon S3](#))
- Dataset features\_prepared ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset train\_final ([Amazon S3](#))



### c) Join

Inner train\_foo - features\_prepared And

Left input	Condition	Right input
Date	=	Date
Store	=	Store

## 5. split\_train\_final

Summarized information about the recipe split\_train\_final.

### a) Status

Characteristics of the recipe split\_train\_final:

- Type: **Split**
- Created by **jayasurya00008** on the 2023-11-30 20:35:42
- Last Modification by **jayasurya00008** on the 2023-11-30 20:36:15

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset train\_final ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset store\_1 ([Amazon S3](#))
- Dataset store\_2 ([Amazon S3](#))
- Dataset store\_3 ([Amazon S3](#))

Split mode: Map values of a single column to the outputs datasets

## 6. compute\_store\_2\_train

Summarized information about the recipe compute\_store\_2\_train.

### a) Status

Characteristics of the recipe compute\_store\_2\_train:

- Type: **Group**
- Created by **jayasurya00008** on the 2023-11-30 20:46:09
- Last Modification by **jayasurya00008** on the 2023-11-30 20:46:11

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_2 ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset store\_2\_train ([Amazon S3](#))

Group keys: Date

## 7. train\_NPTS\_Model

Summarized information about the recipe train\_NPTS\_Model.

### a) Status

Characteristics of the recipe train\_NPTS\_Model:

- Type: **Prediction Training**
- Created by **jayasurya00008** on the 2023-12-01 16:32:11
- Last Modification by **jayasurya00008** on the 2023-12-01 16:32:11

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_2\_train ([Amazon S3](#))

List of the outputs of the recipe:

- Saved Model NPTS\_Model ([NPTS \(store\\_2\\_model\) - v1](#))

## 8. compute\_store\_3\_train

Summarized information about the recipe compute\_store\_3\_train.

### a) Status

Characteristics of the recipe compute\_store\_3\_train:

- Type: **Group**
- Created by **jayasurya00008** on the 2023-11-30 20:46:49
- Last Modification by **jayasurya00008** on the 2023-11-30 20:46:51

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_3 ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset store\_3\_train ([Amazon S3](#))

Group keys: Date

## 9. train\_Ridge\_regression\_automl

Summarized information about the recipe train\_Ridge\_regression\_automl.

### a) Status

Characteristics of the recipe train\_Ridge\_regression\_automl:

- Type: **Prediction Training**
- Created by **jayasurya00008** on the 2023-12-01 16:37:23
- Last Modification by **jayasurya00008** on the 2023-12-01 16:37:23

## b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_3\_train ([Amazon S3](#))

List of the outputs of the recipe:

- Saved Model Ridge\_regression\_automl ([Ridge \(L2\) regression \(auto\\_ml\\_store\\_3\) - v1](#))

## 10. compute\_store\_1\_train

Summarized information about the recipe compute\_store\_1\_train.

### a) Status

Characteristics of the recipe compute\_store\_1\_train:

- Type: **Group**
- Created by **jayasurya00008** on the 2023-11-30 20:41:27
- Last Modification by **jayasurya00008** on the 2023-11-30 20:44:39

### b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_1 ([Amazon S3](#))

List of the outputs of the recipe:

- Dataset store\_1\_train ([Amazon S3](#))

Group keys: Date

## 11. train\_Transformer\_model

Summarized information about the recipe train\_Transformer\_model.

## a) Status

Characteristics of the recipe train\_Transformer\_model:

- Type: **Prediction Training**
- Created by **jayasurya00008** on the 2023-11-30 20:56:51
- Last Modification by **jayasurya00008** on the 2023-11-30 20:56:51

## b) Inputs / Outputs

List of the inputs of the recipe:

- Dataset store\_1\_train ([Amazon S3](#))

List of the outputs of the recipe:

- Saved Model Transformer\_model ([Transformer \(store\\_1\\_model\) - v1](#))

## V. Saved Models

List all the saved models in the flow.

### 1. NPTS\_Model

Summarized information about the saved model NPTS\_Model:.

#### a) Status

Model id: **H43T2jh1**

Active version: **NPTS (store\_2\_model) - v1**

#### b) Related Recipes

List of the recipes related to the saved model.

Parent Recipe

Recipe with NPTS\_Model as an output:

- train\_NPTS\_Model ([Prediction Training](#))

### 2. Ridge\_regression\_automl

Summarized information about the saved model Ridge\_regression\_automl:.

#### a) Status

Model id: **goejdknS**

Active version: **Ridge (L2) regression (auto\_ml\_store\_3) - v1**

#### b) Related Recipes

List of the recipes related to the saved model.

## Parent Recipe

Recipe with Ridge\_regression\_automl as an output:

- train\_Ridge\_regression\_automl ([Prediction Training](#))

### 3. Transformer\_model

Summarized information about the saved model Transformer\_model:.

#### a) Status

Model id: **bJEcHgax**

Active version: **Transformer (store\_1\_model) - v1**

#### b) Related Recipes

List of the recipes related to the saved model.

## Parent Recipe

Recipe with Transformer\_model as an output:

- train\_Transformer\_model ([Prediction Training](#))

## VI. Deployment and Monitoring

### A. Implementation Details

- The flow can be found here: [https://dss-a7fce0f1-51881178-dku.us-east-1.app.dataiku.io/projects/WM\\_SALES\\_FORECAST/](https://dss-a7fce0f1-51881178-dku.us-east-1.app.dataiku.io/projects/WM_SALES_FORECAST/)
- The name of the generated file is: Dataiku Flow Documentation - WM\_SALES\_FORECAST.docx

### B. Version Control

- The flow was executed with the following version of DSS: 12.3.1