Advanced Designer Assessment for jakub.cajzl

short line



|  |  |  |
| --- | --- | --- |
| **Version** | **Author** | **Date** |
| 1.0 | jakub.cajzl | 2024-06-28 07:35:29 |

Table of contents

[I. Project general description 4](#_Toc137204120)

[II. Graphical Representation of the Flow 5](#_Toc137204121)

[III. Datasets 6](#_Toc137204122)

[1. {{$dataset.name}} 6](#_Toc137204123)

[a) Status of the dataset {{$dataset.name}} 6](#_Toc137204124)

[b) Related Recipes 7](#_Toc137204125)

[c) Schema 7](#_Toc137204126)

[IV. Managed folders 9](#_Toc137204127)

[1. {{$folder.name}} 9](#_Toc137204128)

[a) Status 9](#_Toc137204129)

[b) Derived datasets 10](#_Toc137204130)

[c) Related Recipes 10](#_Toc137204131)

[V. Recipes 11](#_Toc137204132)

[1. {{ $recipe.name }} 11](#_Toc137204133)

[a) Status 11](#_Toc137204134)

[b) Inputs / Outputs 11](#_Toc137204135)

[c) Prepare Steps 12](#_Toc137204136)

[d) Join 12](#_Toc137204137)

[e) Data Relationships 13](#_Toc137204138)

[f) Feature Transformations 14](#_Toc137204139)

[g) Code 14](#_Toc137204140)

[h) Windows 15](#_Toc137204141)

[i) Sorted Columns 15](#_Toc137204142)

[j) Top rows and sorting direction 15](#_Toc137204143)

[k) Parameters 16](#_Toc137204144)

[VI. Saved Models 17](#_Toc137204145)

[1. {{ $model.name }} 17](#_Toc137204146)

[a) Status 17](#_Toc137204147)

[b) Related Recipes 17](#_Toc137204148)

[VII. Model Evaluation Stores 19](#_Toc137204149)

[1. {{ $mes.name }} 19](#_Toc137204150)

[a) Related Recipes 19](#_Toc137204151)

[VIII. Labeling Tasks 21](#_Toc137204152)

[1. {{ $labeling\_task.name }} 21](#_Toc137204153)

[a) Status 21](#_Toc137204154)

[b) Inputs / Outputs 21](#_Toc137204155)

[IX. Deployment and Monitoring 23](#_Toc137204156)

[A. Implementation Details 23](#_Toc137204157)

[B. Version Control 23](#_Toc137204158)

# Project general description

Project references for the created flow:

* Project name: **Advanced Designer Assessment for jakub.cajzl**
* Project key: **DKU\_EXAM\_ADV\_DESIGNER**
* Project short description:

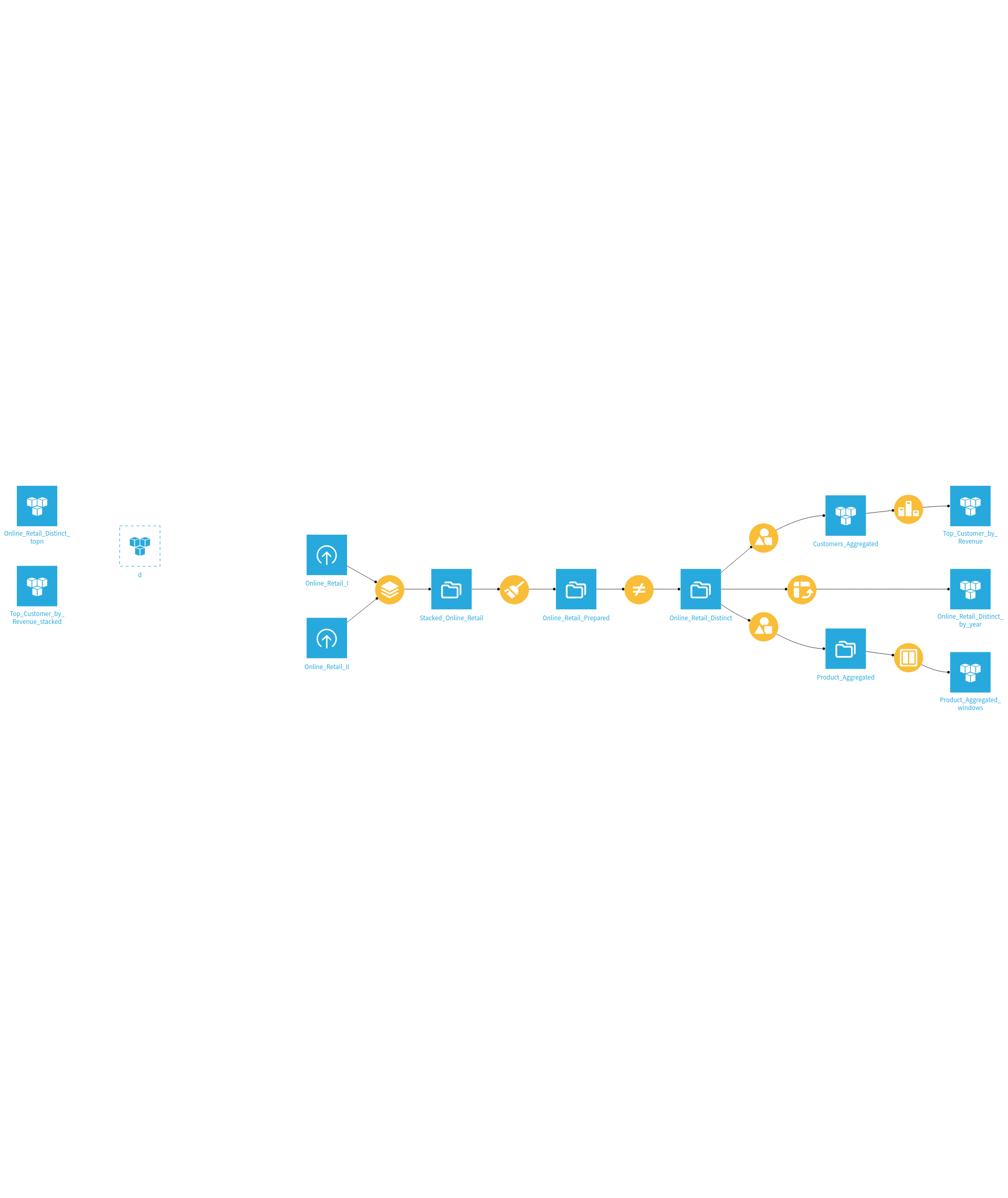
This project provides a starting point for the [Advanced Designer Assessment](https://academy.dataiku.com/advanced-designer-certificate).

* Description:

Uploaded data originally acquired from the UCI ML Repositories [Online Retail](https://archive.ics.uci.edu/ml/datasets/Online+Retail) and [Online Retail II Data Set](https://archive.ics.uci.edu/ml/datasets/Online+Retail+II)

* Status: **Sandbox**
* Current git branch: **master**
* Tags: **academy, exam, tutorial, downloads, section:Advanced Designer**
* Created by admin (deleted) on the 2020-11-20 09:26:53
* Last Modification by api:G3846H2B5H70ZESZ (deleted) on the 2024-06-20 14:41:52

# Graphical Representation of the Flow



# Datasets

List all the datasets in the flow.

### Online\_Retail\_I

Summarized information about your dataset Online\_Retail\_I:

* Tags**: Skeleton Project**

#### Status of the dataset Online\_Retail\_I

Characteristics of the dataset Online\_Retail\_I:

* Type: **Uploaded Files**
* Connection:
* Total Size: **Not computed**
* Record Count: **Not computed**
* Created by admin (deleted) on the 2020-12-02 09:10:42
* Last Modification by admin (deleted) on the 2020-12-02 09:32:18

#### Related Recipes

The dataset Online\_Retail\_I relates to the following recipes.

Successor Recipes

List of recipes with Online\_Retail\_I as an input:

* compute\_Online\_Retail\_stacked (Stack)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****InvoiceNo**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****StockCode**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Quantity**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****InvoiceDate**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****UnitPrice**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****CustomerID**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Country**** | string |  |

### Online\_Retail\_II

Summarized information about your dataset Online\_Retail\_II:

* Tags**: Skeleton Project**

#### Status of the dataset Online\_Retail\_II

Characteristics of the dataset Online\_Retail\_II:

* Type: **Uploaded Files**
* Connection:
* Total Size: **Not computed**
* Record Count: **Not computed**
* Created by admin (deleted) on the 2020-12-02 09:11:14
* Last Modification by admin (deleted) on the 2020-12-02 09:32:18

#### Related Recipes

The dataset Online\_Retail\_II relates to the following recipes.

Successor Recipes

List of recipes with Online\_Retail\_II as an input:

* compute\_Online\_Retail\_stacked (Stack)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****Invoice**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****StockCode**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Quantity**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****InvoiceDate**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Price**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Customer ID**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Country**** | string |  |

### Stacked\_Online\_Retail

Summarized information about your dataset Stacked\_Online\_Retail:

* Tags**: Skeleton Project**

#### Status of the dataset Stacked\_Online\_Retail

Characteristics of the dataset Stacked\_Online\_Retail:

* Type: **Server's Filesystem**
* Connection: **filesystem\_managed**
* Total Size: **14.1 MB**
* Record Count: **1,067,370**
* Created by admin (deleted) on the 2020-11-20 14:30:17
* Last Modification by admin (deleted) on the 2020-11-25 11:19:34

#### Related Recipes

The dataset Stacked\_Online\_Retail relates to the following recipes.

Parent Recipe

List of recipes with Stacked\_Online\_Retail as an output:

* compute\_Online\_Retail\_stacked (Stack)

Successor Recipes

List of recipes with Stacked\_Online\_Retail as an input:

* compute\_Online\_Retail\_Prepared (Prepare)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****InvoiceNo**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****StockCode**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Quantity**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****InvoiceDate**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****UnitPrice**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****CustomerID**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Country**** | string |  |

### Online\_Retail\_Prepared

Summarized information about your dataset Online\_Retail\_Prepared:

* Tags**: Part Skeleton Project**

#### Status of the dataset Online\_Retail\_Prepared

Characteristics of the dataset Online\_Retail\_Prepared:

* Type: **Server's Filesystem**
* Connection: **filesystem\_managed**
* Total Size: **9 MB**
* Record Count: **725,284**
* Created by admin (deleted) on the 2020-11-20 14:39:17
* Last Modification by jakub.cajzl on the 2024-06-20 14:00:11

#### Related Recipes

The dataset Online\_Retail\_Prepared relates to the following recipes.

Parent Recipe

List of recipes with Online\_Retail\_Prepared as an output:

* compute\_Online\_Retail\_Prepared (Prepare)

Successor Recipes

List of recipes with Online\_Retail\_Prepared as an input:

* compute\_Online\_Retail\_Prepared\_distinct (Distinct)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****customerid**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****stock\_code**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****year**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****quantity**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue**** | double |  |

### Online\_Retail\_Distinct

Summarized information about your dataset Online\_Retail\_Distinct:

* Tags**: Skeleton Project**

#### Status of the dataset Online\_Retail\_Distinct

Characteristics of the dataset Online\_Retail\_Distinct:

* Type: **Server's Filesystem**
* Connection: **filesystem\_managed**
* Total Size: **8.2 MB**
* Record Count: **694,384**
* Created by admin (deleted) on the 2020-11-23 07:48:40
* Last Modification by jakub.cajzl on the 2024-06-21 07:58:22

#### Related Recipes

The dataset Online\_Retail\_Distinct relates to the following recipes.

Parent Recipe

List of recipes with Online\_Retail\_Distinct as an output:

* compute\_Online\_Retail\_Prepared\_distinct (Distinct)

Successor Recipes

List of recipes with Online\_Retail\_Distinct as an input:

* compute\_Online\_Retail\_Distinct\_by\_customerid (Group)
* compute\_Online\_Retail\_Distinct\_by\_year (Pivot)
* compute\_Product\_Aggregated\_1 (Group)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****customerid**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****stock\_code**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****year**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****quantity**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue**** | double |  |

### Product\_Aggregated

Summarized information about your dataset Product\_Aggregated:

* Tags**: Product-Centric-Flow, Skeleton Project**

#### Status of the dataset Product\_Aggregated

Characteristics of the dataset Product\_Aggregated:

* Type: **Server's Filesystem**
* Connection: **filesystem\_managed**
* Total Size: **1.9 KB**
* Record Count: **101**
* Created by admin (deleted) on the 2020-11-27 13:48:59
* Last Modification by jakub.cajzl on the 2024-06-21 12:09:35

#### Related Recipes

The dataset Product\_Aggregated relates to the following recipes.

Parent Recipe

List of recipes with Product\_Aggregated as an output:

* compute\_Product\_Aggregated\_1 (Group)

Successor Recipes

List of recipes with Product\_Aggregated as an input:

* compute\_Product\_Aggregated\_windows (Window)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****stock\_code**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****quantity\_avg**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****total\_quantity**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****count**** | bigint |  |

### Product\_Aggregated\_windows

Summarized information about your dataset Product\_Aggregated\_windows:

#### Status of the dataset Product\_Aggregated\_windows

Characteristics of the dataset Product\_Aggregated\_windows:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **4.7 KB**
* Record Count: **101**
* Created by jakub.cajzl on the 2024-06-21 08:37:16
* Last Modification by jakub.cajzl on the 2024-06-21 12:09:59

#### Related Recipes

The dataset Product\_Aggregated\_windows relates to the following recipes.

Parent Recipe

List of recipes with Product\_Aggregated\_windows as an output:

* compute\_Product\_Aggregated\_windows (Window)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****stock\_code**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****quantity\_avg**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****total\_quantity**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****count**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****days\_since\_last\_purchase**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_quantity\_avg**** | double |  |

### Online\_Retail\_Distinct\_by\_year

Summarized information about your dataset Online\_Retail\_Distinct\_by\_year:

#### Status of the dataset Online\_Retail\_Distinct\_by\_year

Characteristics of the dataset Online\_Retail\_Distinct\_by\_year:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **4.8 KB**
* Record Count: **104**
* Created by jakub.cajzl on the 2024-06-21 07:51:46
* Last Modification by jakub.cajzl on the 2024-06-21 07:54:03

#### Related Recipes

The dataset Online\_Retail\_Distinct\_by\_year relates to the following recipes.

Parent Recipe

List of recipes with Online\_Retail\_Distinct\_by\_year as an output:

* compute\_Online\_Retail\_Distinct\_by\_year (Pivot)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****2009\_quantity\_avg**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2009\_total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2009\_count**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****2010\_quantity\_avg**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2010\_total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2010\_count**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****2011\_quantity\_avg**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2011\_total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****2011\_count**** | bigint |  |

### Customers\_Aggregated

Summarized information about your dataset Customers\_Aggregated:

#### Status of the dataset Customers\_Aggregated

Characteristics of the dataset Customers\_Aggregated:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **97 KB**
* Record Count: **8,604**
* Created by jakub.cajzl on the 2024-06-21 08:18:05
* Last Modification by jakub.cajzl on the 2024-06-21 08:22:38

#### Related Recipes

The dataset Customers\_Aggregated relates to the following recipes.

Parent Recipe

List of recipes with Customers\_Aggregated as an output:

* compute\_Online\_Retail\_Distinct\_by\_customerid (Group)

Successor Recipes

List of recipes with Customers\_Aggregated as an input:

* compute\_Top\_Customer\_by\_Revenue (Top N)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****customerid**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****year**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****count**** | bigint |  |

### Top\_Customer\_by\_Revenue

Summarized information about your dataset Top\_Customer\_by\_Revenue:

#### Status of the dataset Top\_Customer\_by\_Revenue

Characteristics of the dataset Top\_Customer\_by\_Revenue:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **1.4 KB**
* Record Count: **15**
* Created by jakub.cajzl on the 2024-06-21 08:24:47
* Last Modification by jakub.cajzl on the 2024-06-21 08:26:11

#### Related Recipes

The dataset Top\_Customer\_by\_Revenue relates to the following recipes.

Parent Recipe

List of recipes with Top\_Customer\_by\_Revenue as an output:

* compute\_Top\_Customer\_by\_Revenue (Top N)

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****customerid**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****year**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue\_sum**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****count**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****\_rank**** | bigint |  |

### Top\_Customer\_by\_Revenue\_stacked

Summarized information about your dataset Top\_Customer\_by\_Revenue\_stacked:

#### Status of the dataset Top\_Customer\_by\_Revenue\_stacked

Characteristics of the dataset Top\_Customer\_by\_Revenue\_stacked:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **Not computed**
* Record Count: **Not computed**
* Created by jakub.cajzl on the 2024-06-21 12:40:26
* Last Modification by jakub.cajzl on the 2024-06-21 12:40:26

#### Related Recipes

The dataset Top\_Customer\_by\_Revenue\_stacked relates to the following recipes.

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****InvoiceNo**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****StockCode**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Quantity**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****InvoiceDate**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****UnitPrice**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****CustomerID**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Country**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Invoice**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Price**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****Customer ID**** | string |  |

### d

Summarized information about your dataset d:

#### Status of the dataset d

Characteristics of the dataset d:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **Not computed**
* Record Count: **Not computed**
* Created by jakub.cajzl on the 2024-06-28 07:26:23
* Last Modification by jakub.cajzl on the 2024-06-28 07:26:23

#### Related Recipes

The dataset d relates to the following recipes.

#### Schema

This dataset doesn’t have any schema

### Online\_Retail\_Distinct\_topn

Summarized information about your dataset Online\_Retail\_Distinct\_topn:

#### Status of the dataset Online\_Retail\_Distinct\_topn

Characteristics of the dataset Online\_Retail\_Distinct\_topn:

* Type: **Amazon S3**
* Connection: **dataiku-managed-storage**
* Total Size: **442.2 KB**
* Record Count: **15**
* Created by jakub.cajzl on the 2024-06-21 08:03:37
* Last Modification by jakub.cajzl on the 2024-06-21 08:09:16

#### Related Recipes

The dataset Online\_Retail\_Distinct\_topn relates to the following recipes.

#### Schema

Schema details of the dataset.

|  |  |  |
| --- | --- | --- |
| ****Column Name**** | Column Type | Description |

|  |  |  |
| --- | --- | --- |
| ****customerid**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****stock\_code**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****description**** | string |  |

|  |  |  |
| --- | --- | --- |
| ****week\_of**** | date |  |

|  |  |  |
| --- | --- | --- |
| ****year**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****quantity**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****total\_revenue**** | double |  |

|  |  |  |
| --- | --- | --- |
| ****\_duplicate\_count**** | bigint |  |

|  |  |  |
| --- | --- | --- |
| ****\_rank**** | bigint |  |

# Recipes

List all the recipes in the flow.

### compute\_Online\_Retail\_stacked

Summarized information about the recipe compute\_Online\_Retail\_stacked.

* Tags**: Skeleton Project**

#### Status

Characteristics of the recipe compute\_Online\_Retail\_stacked:

* Type: **Stack**
* Created by admin (deleted) on the 2020-11-20 14:30:17
* Last Modification by admin (deleted) on the 2020-12-10 19:17:10

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Online\_Retail\_I (Uploaded Files)
* Dataset Online\_Retail\_II (Uploaded Files)

List of the outputs of the recipe:

* Dataset Stacked\_Online\_Retail (Server's Filesystem)

Stacking mode: Using column order

### compute\_Online\_Retail\_Prepared

Summarized information about the recipe compute\_Online\_Retail\_Prepared.

* Tags**: Part Skeleton Project**

#### Status

Characteristics of the recipe compute\_Online\_Retail\_Prepared:

* Type: **Prepare**
* Created by admin (deleted) on the 2020-11-20 14:39:17
* Last Modification by jakub.cajzl on the 2024-06-21 12:13:27

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Stacked\_Online\_Retail (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Online\_Retail\_Prepared (Server's Filesystem)

#### Prepare Steps

|  |  |
| --- | --- |
| Type | Comment |

|  |  |
| --- | --- |
| DateParser |  |

|  |  |
| --- | --- |
| CreateColumnWithGREL | Truncating Invoice Date to the week level |

|  |  |
| --- | --- |
| DateComponentsExtractor |  |

|  |  |
| --- | --- |
| FilterOnBadType | Invoices starting with 'C' indicate cancelled orders |

|  |  |
| --- | --- |
| FilterOnNumericalRange |  |

|  |  |
| --- | --- |
| FilterOnValue |  |

|  |  |
| --- | --- |
| RemoveRowsOnEmpty |  |

|  |  |
| --- | --- |
| TextSimplifierProcessor |  |

|  |  |
| --- | --- |
| StringTransformer |  |

|  |  |
| --- | --- |
| StringTransformer |  |

|  |  |
| --- | --- |
| ColumnRenamer |  |

|  |  |
| --- | --- |
| ColumnReorder |  |

|  |  |
| --- | --- |
| CreateColumnWithGREL |  |

|  |  |
| --- | --- |
| ColumnReorder |  |

|  |  |
| --- | --- |
| ColumnsSelector |  |

|  |  |
| --- | --- |
| FilterOnValue |  |

|  |  |
| --- | --- |
| CreateColumnWithGREL |  |

### compute\_Online\_Retail\_Prepared\_distinct

Summarized information about the recipe compute\_Online\_Retail\_Prepared\_distinct.

* Tags**: Skeleton Project**

#### Status

Characteristics of the recipe compute\_Online\_Retail\_Prepared\_distinct:

* Type: **Distinct**
* Created by admin (deleted) on the 2020-11-23 07:48:40
* Last Modification by jakub.cajzl on the 2024-06-21 07:58:22

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Online\_Retail\_Prepared (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Online\_Retail\_Distinct (Server's Filesystem)

Distinct keys: customerid, stock\_code, description, week\_of, year, quantity, total\_revenue

### compute\_Product\_Aggregated\_1

Summarized information about the recipe compute\_Product\_Aggregated\_1.

* Tags**: Part Skeleton Project, Product-Centric-Flow**

#### Status

Characteristics of the recipe compute\_Product\_Aggregated\_1:

* Type: **Group**
* Created by admin (deleted) on the 2020-11-27 13:48:59
* Last Modification by jakub.cajzl on the 2024-06-21 12:09:35

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Online\_Retail\_Distinct (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Product\_Aggregated (Server's Filesystem)

Group keys: stock\_code, week\_of

### compute\_Product\_Aggregated\_windows

Summarized information about the recipe compute\_Product\_Aggregated\_windows.

#### Status

Characteristics of the recipe compute\_Product\_Aggregated\_windows:

* Type: **Window**
* Created by jakub.cajzl on the 2024-06-21 08:37:16
* Last Modification by jakub.cajzl on the 2024-06-21 08:43:32

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Product\_Aggregated (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Product\_Aggregated\_windows (Amazon S3)

#### Windows

|  |  |
| --- | --- |
| Window name | Partitioning columns |

|  |  |
| --- | --- |
|  | stock\_code |

### compute\_Online\_Retail\_Distinct\_by\_year

Summarized information about the recipe compute\_Online\_Retail\_Distinct\_by\_year.

#### Status

Characteristics of the recipe compute\_Online\_Retail\_Distinct\_by\_year:

* Type: **Pivot**
* Created by jakub.cajzl on the 2024-06-21 07:51:46
* Last Modification by jakub.cajzl on the 2024-06-21 07:53:48

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Online\_Retail\_Distinct (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Online\_Retail\_Distinct\_by\_year (Amazon S3)

#### Parameters

Row identifier: week\_of

Pivots: year

Populate content: Count of records, quantity, total\_revenue

### compute\_Online\_Retail\_Distinct\_by\_customerid

Summarized information about the recipe compute\_Online\_Retail\_Distinct\_by\_customerid.

#### Status

Characteristics of the recipe compute\_Online\_Retail\_Distinct\_by\_customerid:

* Type: **Group**
* Created by jakub.cajzl on the 2024-06-21 08:18:05
* Last Modification by jakub.cajzl on the 2024-06-21 12:32:15

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Online\_Retail\_Distinct (Server's Filesystem)

List of the outputs of the recipe:

* Dataset Customers\_Aggregated (Amazon S3)

Group keys: customerid, year

### compute\_Top\_Customer\_by\_Revenue

Summarized information about the recipe compute\_Top\_Customer\_by\_Revenue.

#### Status

Characteristics of the recipe compute\_Top\_Customer\_by\_Revenue:

* Type: **Top N**
* Created by jakub.cajzl on the 2024-06-21 08:24:47
* Last Modification by jakub.cajzl on the 2024-06-21 12:33:40

#### Inputs / Outputs

List of the inputs of the recipe:

* Dataset Customers\_Aggregated (Amazon S3)

List of the outputs of the recipe:

* Dataset Top\_Customer\_by\_Revenue (Amazon S3)

#### Top rows and sorting direction

Top rows: 5

Bottom rows: 0

|  |  |
| --- | --- |
| Column name | Sorting direction |

|  |  |
| --- | --- |
| total\_revenue\_sum | Descending |

# Deployment and Monitoring

## Implementation Details

* The flow can be found here: https://dss-b2df749f-ec12f1b3-dku.eu-west-2.app.dataiku.io/projects/DKU\_EXAM\_ADV\_DESIGNER/
* The name of the generated file is: Dataiku Flow Documentation - DKU\_EXAM\_ADV\_DESIGNER.docx

## Version Control

* The flow was executed with the following version of DSS: 12.6.2