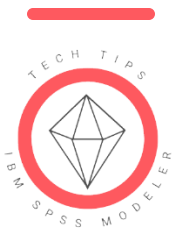




---

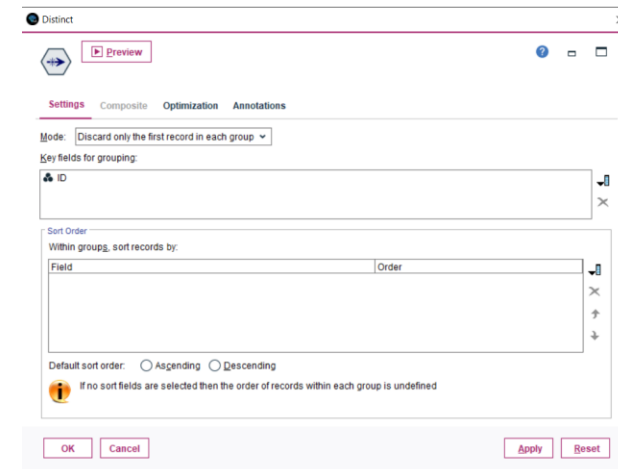
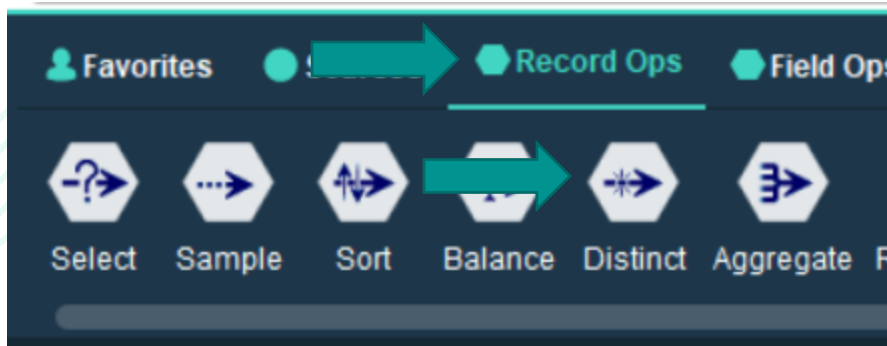
## Tech Tips - Find Duplicate Records in IBM SPSS Modeler

## Tech Tips - Find Duplicate Records in IBM SPSS Modeler



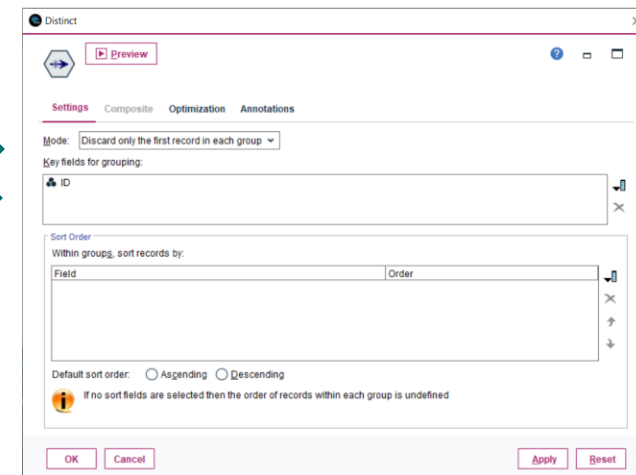
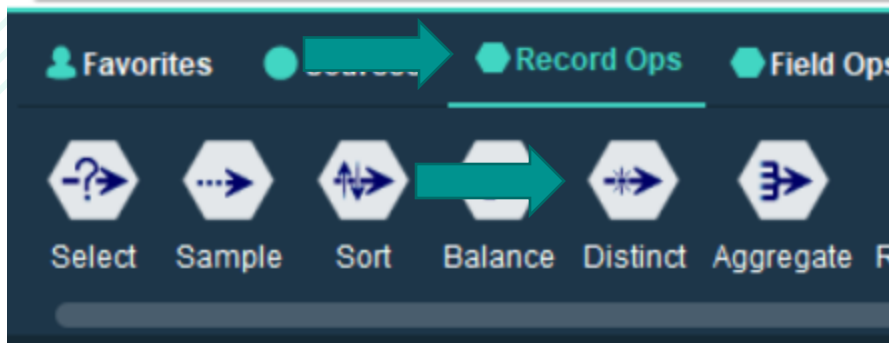
# Tech Tips - Find Duplicate Records

- Here's a quick tip to find duplicate records in IBM SPSS Modeler.
- The **Distinct** node makes identifying duplicates quick and easy and is located on the **Record Ops** palette.



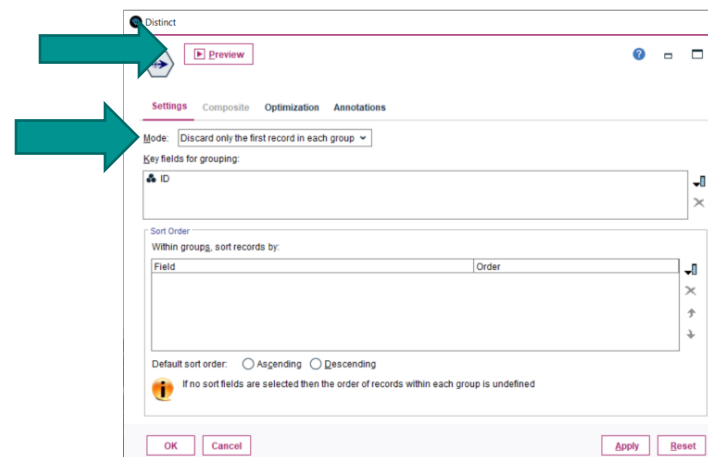
# Tech Tips - Find Duplicate Records

- To identify duplicates go to the **Record Ops** palette. Select the **Distinct** node and drag it onto the stream canvas. You can also double click the node to drop it onto the stream canvas. Once it is on the canvas you can connect it to your stream.
- Double click to open the node. We want to identify records that are duplicates so that we can discard them. We will set the **Mode** to **Discard only the first record in each group**. Use the field chooser button to select the field. Here we are looking for to identify duplicate records for customers.



# Tech Tips - Find Duplicate Records

- On the Distinct node, the **Include** and **Discard** options control whether the first distinct record is passed (**Include**) or all but the first distinct record are passed (**Discard**).
- To remove duplicates in a database set the **Mode** to **Include only the first record in each group**; to identify duplicates set the **Mode** to **Discard only the first record in each group**. Fields that provide the basis for identification of duplicates are selected in the **Key fields for grouping** list. To check your data, click on the **Preview** button.



# SPSS Tech Tips

For more **Tech Tips**, go to the Tech Tips section within the Learning Hub.

## Contact us



[SPSSAdmin@version1.com](mailto:SPSSAdmin@version1.com)



+ 44 (0) 203 859 4790  
+ 353 (0) 1 865 7800



[www.spssanalyticspartner.com](http://www.spssanalyticspartner.com)



**Thank you.**