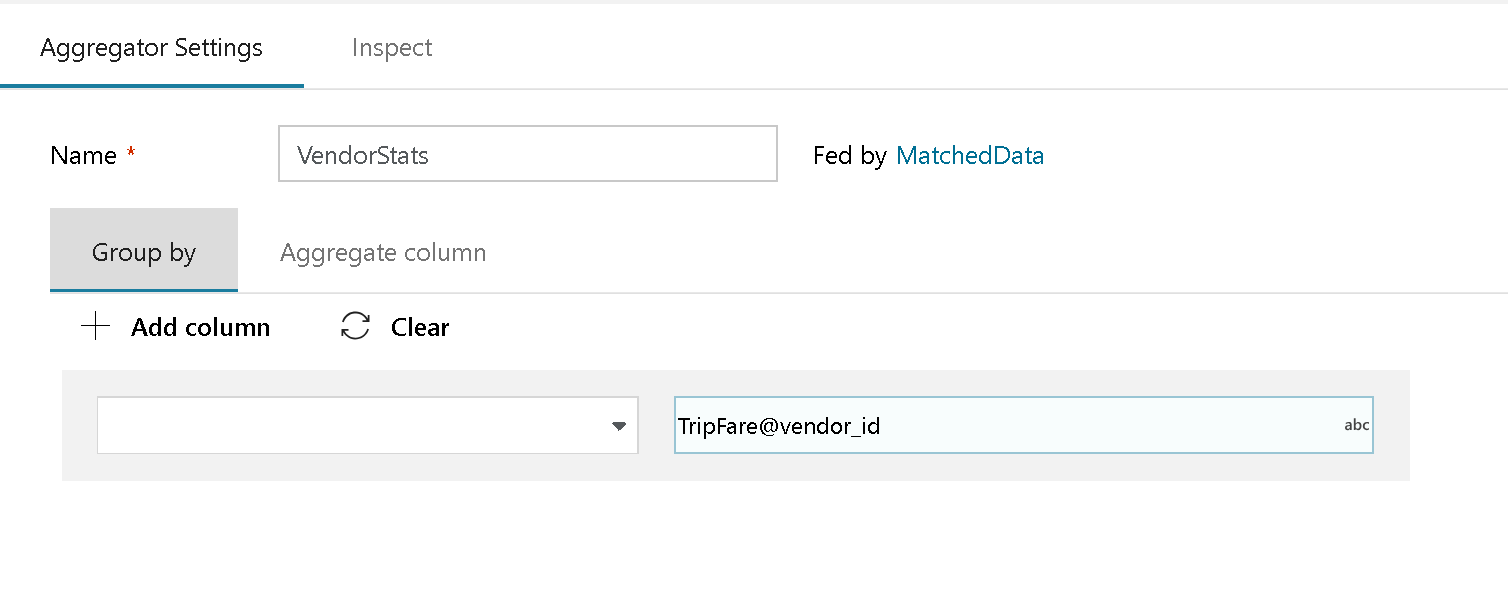
# Azure Data Factory Data Flow Transformations

## Aggregate

The Aggregate transformation is where you will define aggregations of columns in your data streams. In the Expression Builder, you can define different types of aggregations (i.e. SUM, MIN, MAX, COUNT, etc …) and create a new field in your output that includes these aggregations with optional group-by fields.



Agg Transformation options

*NOTE: Aggregate transforms will only output the columns used in the aggregation. I.e. only the fields used in group-by and the aggregated fields will be passed on to the next transformation in your data flow. If you wish to include the previous columns in your flow, use a New Branch from the previous step and use the self-join pattern to connect the flow with the original metadata.*

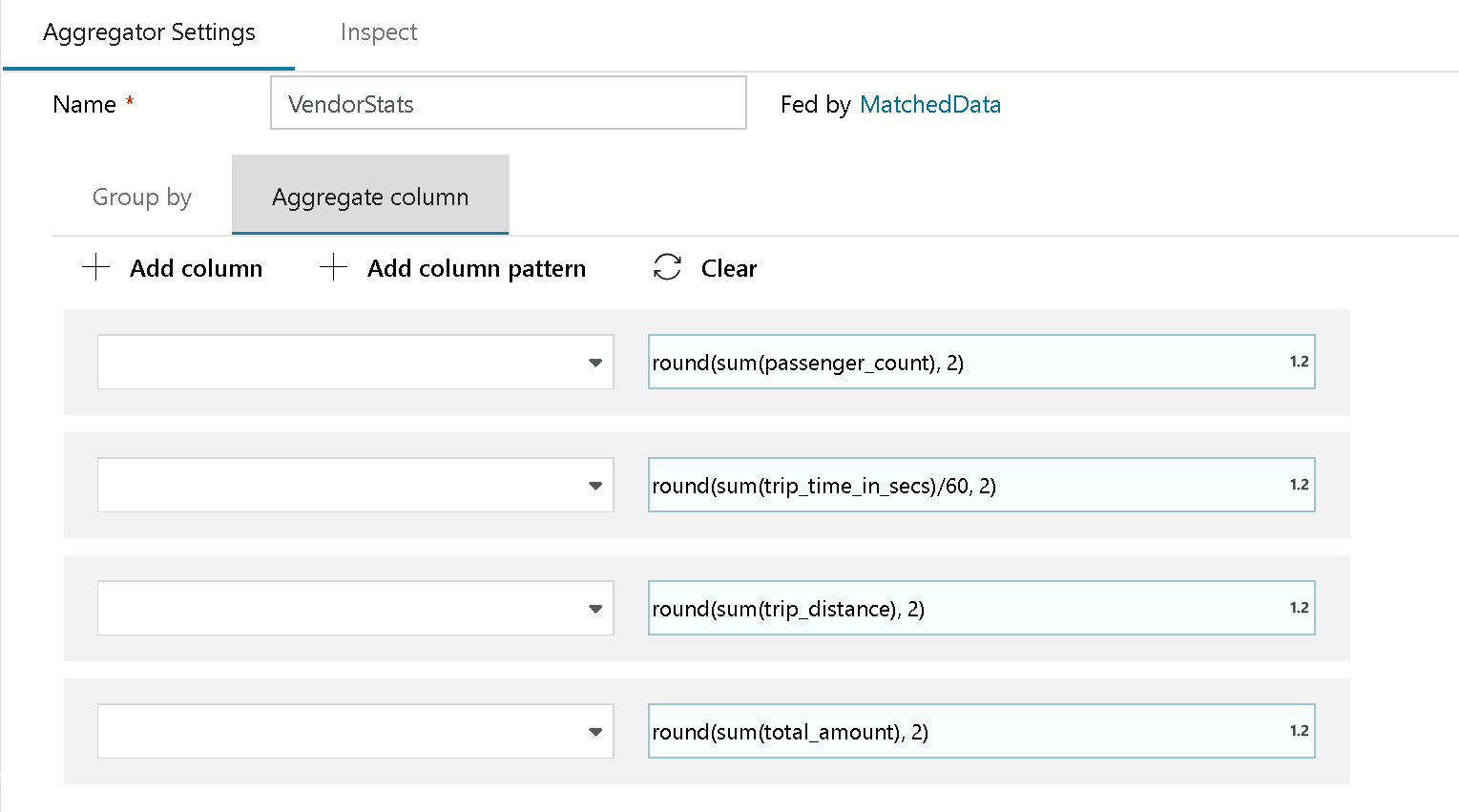
### Group By

(Optional) Choose a Group-by clause for your aggregation and use either the name of an existing column or a new name. Use “Add Column” add more group-by clauses and click on the text box next to the column name to launch the Expression Builder to either select just an existing column, combination of columns or expressions for your grouping.

### The Aggregate Column tab

(Required) Choose the Aggregate Column tab to build the aggregation expressions. You can either choose an existing column to overwrite the value with the aggregation, or create a new field with the new name for the aggregation. The expression that you wish to use for the aggregation will be entered in the right-hand box next to the column name selector. Clicking on that text box will open up the Expression Builder.

Use the ADF Data Flow Expression Language to describe the column transformations in the Expression Builder: https://aka.ms/dataflowexpressions.



Agg Transformation options

### Data Preview in Expression Builder

Note that in Debug mode, expression builder cannot produce data previews with Aggregate functions. You must view data previews from the Aggregate transformation, outside of the expression builder.

### Use Aggregates for Data Deduping

You can use the grouping function in the Aggregate Transformation to group data by an attribute that can help identify and remove duplicates from your data sources: [ADF Data Flow: Deduplication of your Data](https://www.youtube.com/watch?v=OLenvYwg__I).