

Data Preparation- 6.0



Conte	ents		
1. Intr	oduct	ion	4
1.1.	Intro	oducing Data Preparation	4
1.2.	Doc	ument History	4
1.3.	Ove	rview	4
1.4.	Targ	et Audience	4
2. Sup	porte	d Web Browsers	4
2.1.	Uns	upported Browsers	5
3. Get	ting S	tarted with BDB Data Preparation	5
4. Dat	a Pre	paration Landing Page	7
4.1.	Prep	parations	7
4.1.	.1.	Importing a Preparation	8
4.2.	Data	asets	9
4.2.	.1.	Adding a new Dataset	9
5. Dat	a Gric	I	10
5.1.	Data	a Grid Header	11
5.1.	.1.	Data Types	12
5.2.	Pane	el to List the Selected Filters	12
5.3.	Data	a Quality Bar in the Grid	13
5.4.	Pagi	nation	14
6. Sun	nmary	/ Pane	15
6.1.	Chai	rts	15
6.2.	Info	: Value/Statistics	15
6.3.	Patt	ern	17
6.4.	Tran	isforms	17
6.4.	.1.	Advanced	17
6.4.	.2.	Anonymization	21
6.4.	.3.	Columns	24
6.4.	.4.	Conversions	33
6.4.	.5.	Data Cleansing	34
6.4.	.6.	Dates	43
6.4.	.7.	Integer	48
6.4.	.8.	ML	49
6.4.	.9.	Numbers	49
6.4.	.10.	String	51



	6.5.	Steps	54
	6.5.3	1. Deleting or Undeleting Steps	55
7.	Navi	igation Pane	57
	7.1.	Export Steps to a Data Store Meta Data	57
	7.2.	Export Steps to Pipeline	59
	7.3.	Other options in the Navigation Pane	60
8.	Sign	ing Out	60
	8.1.	Forgot Password Option	62
	8.2.	Force Login	63



1. Introduction

1.1. Introducing Data Preparation

The Data Preparation module can turn any Business data into a cost-effective and custom-made experience. The Data Analysts can instantly detect anomalous records (rows with invalid or empty values) and purge the unwanted data sets in a few clicks using Machine-Learning based smart techniques and sampling. The users can identify errors and apply changes to data set from any source and export the analysis-ready data in minutes. Automated detection of groups and categories in your data can be viewed through a frequency table. The user can filter the group in a single click and transform data matching the filter conditions and get intelligent Data Transformation suggestions based on data type and quality.

1.2. Document History

Product Version	Date (Release Date)	Description
Data Preparation 4.0	December 31 st , 2018	First Release of the document
Data Preparation 4.2	March 25 th , 2019	Updated document
Data Preparation 4.3	April 24 th , 2019	Updated document
Data Preparation 4.4	June 7 th , 2019	Updated document
Data Preparation 4.5	August 5 th , 2019	Updated document
Data Preparation 4.6	November 15 th , 2019	Updated document
Data Preparation 5.2.0	August 21 st , 2020	Updated document
Data Preparation 6.0	February 26 th , 2021	Updated document

Note: The Product Version column displays Release Version of the BDB Platform.

1.3. Overview

This guide covers:

- Explanation and usage of all the Data Preparation options
- Explanation and usage of the Transforms
- Integration with Data Pipeline

1.4. Target Audience

The document is targeted to the following audience:

- Data Engineers
- Citizen Data Scientists

2. Supported Web Browsers

The BDB Platform is a web browser-based application. The users can run the BDB Platform and its various plugins on the below given versions of the browsers:

Google Chrome	Latest Version (recommended web browser)
Mozilla Firefox/ Firefox ESR	Latest Version
Microsoft Edge	Latest Version
Apple Safari	10



The supported browser versions are driven by the capabilities the UI employs and the dependencies it uses. UI features will be developed and tested against the supported browsers.

2.1. Unsupported Browsers

While the UI may run successfully in unsupported browsers, it is not actively tested against them. Additionally, the UI is designed as a desktop experience and is not currently supported in mobile browsers.

3. Getting Started with BDB Data Preparation

This section explains how to access the BDB Platform and a variety of plugins that it offers:

- i) Open BDB Enterprise Platform Link: https://app.bdb.ai
- ii) Enter your credentials.
- iii) Select an Auth Type from the drop-down menu.
- iv) Click the 'Sign In' option.



v) The Platform homepage opens.

My Documents			
Welcome to BDB Decision platform	Dashboard Designer	Business Story	Data Science Workbench
BDB is a complete decision platform for all your business needs. Drive from data to dynamic visuals and derive an actionable insight into your business data. Avail 360' view of your	Design, save and publish a splendid governed dashboards. Display relevant KPIs through comprehensive and stunning visual reports to attain your business objectives.	Go beyond the classic BI with our ground-breaking self-service BI tool. Gain pertinent insights into your business by creating wide-ranging views on your own without external IT help.	Let the power of advanced statistical analysis and machine learning plan your next business move. Access and apply accurate and customizable Predictive model to maximize future opportunities.
business by assembling, processing, and analyzing the acquired data. Access incomparable analytics at	Data Center	etl	Data Preparation
any time from anywhere on any device.	Supports a wide range of Data sources starting from the spreadsheets in your system to a cloud-based database. Establish connections to these data sources and build Data Sets or Data Stores	A self-driven Data Wrangling tool to extract data from diverse sources, including the merged data. Enforce data quality and consistency standards to deliver the output in a presentation-ready format.	Experience a secure yet self-driven mode of data preparation. Streamline the entire process of dealing with retrievable business data empowering the business users to decide with unprecedente

Copyright © 2015-2021 BDB

www.bdb.ai



Note:

- **a.** The above screen opens only for those newly created users who have not yet created any document using the BDB Platform.
- **b.** If the user has created some documents previously, then the Platform homepage opens displaying the '**My Documents**' page by default.
- vi) Click the 'Apps' 🎫 icon.
- vii) All the available plugin applications get displayed.
- viii) Select the 'Data Preparation' plugin.



- ix) The Data Preparation landing page opens.
- x) The major Data Preparation modules get displayed on the landing page:
 - a. Home (Default module)
 - b. Preparations
 - c. Datasets

Data Preparation			Take a Tou
Home			Ret
Preparations			
	Conversant Interface	Data Profiling	Instant Filter
Datasets	An enterprise-grade paginated grid UI to display data sample directly to the user. Start exploring your source data from the very first glance of it.	Shape the existing source data by recognizing relevant patterns and identifying data anomalies discovering the metadata of it.	Filter groups in a single click and transform data matching filter conditions. Get Intelligent Data Transformation suggestions based on data type an quality or undo the applied steps just in one click.
	9	E C	•••
	Vigorous Data Grounding	Avail a Complete Picture of your Data	Export and Share Preparations
	Empower users to cleanse data effortlessly with the self-service Data Preparation tool. Transform your business data into trusted insights by interacting with it at each steo.	Access a quality bar indicating profiles of each column and your entire data preparation process side by side. Pick which view to pay attention based on your need to interact with data.	Export preparations to our ETL and run at big data scale. Export to our Data Pipeline to productionised automate the data unification for real-time or batch data

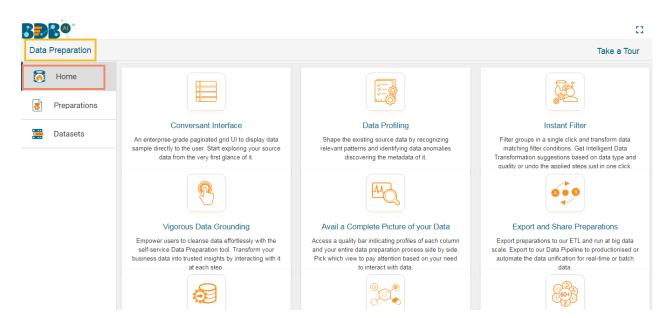


- xi) The user can start a guided tour of the plugin by clicking on the 'Take a Tour' option.
- xii) Click the 'Next' option.

This document aims to describe all the significant components and the related workflows at the detail.

4. Data Preparation Landing Page

The landing page of the data preparation has three menus: 1) Home, 2) Preparations, and 3) Datasets. The 'Home' page opens by default while selecting 'Data Preparation' plugin from the Apps menu.



The user can start the data preparation process by uploading a dataset, and the newly created preparation gets saved under the '**Preparations**' tab.

4.1. Preparations

The '**Preparations**' tab lists all the available preparations displaying Name, Author, when it was created, when it was last modified and using which data`1set it was created.

Data Preparation											Take a Tour
🔂 Home											☐ Import Preparatio
Preparations	Type here to search a Prep	paratio	n								Q
-	Name							Author	Created	Modified ▲	Dataset
Datasets	New Microsoft Excel Worl		Ľ	/	6	Ŧ		William Martin	8 months ago	8 months ago	New Microsoft Excel Wor
	Hiring Data(CSV)_1	7	Ľ	/	6	Ŧ	I	William Martin	8 months ago	8 months ago	Hiring Data(CSV)
	Sample Data_2	7	Ľ	/	6	Ŧ	I	William Martin	8 months ago	8 months ago	Sample Data
	Hiring Data(CSV)_3	7	Ľ	/	6	Ŧ		William Martin	8 months ago	8 months ago	Hiring Data(CSV)
	Hiring Data(CSV)_4	7	K	/	Ē	ŧ		William Martin	8 months ago	8 months ago	Hiring Data(CSV)



The user can continue adding more steps to the existing preparations. The user can import an existing preparation using the '**Import Preparation**' option.

4.1.1. Importing a Preparation

This feature can be used to apply a set of cleansing steps on a dataset with similar metadata.

- i) Navigate to the 'Preparations' list.
- ii) Click the 'Import Preparation' option.

	Preparation					Take a Tour
<u> </u>	Home					2 T Import Preparatio
	Preparations	Type here to search a Preparation				a
		Name	Author	Created	Modified	Dataset
	Datasets	New Microsoft Excel Works 7 V / D ±	William Martin	8 months ago	8 months ago	New Microsoft Excel Wor
		Hiring Data(CSV)_1	William Martin	8 months ago	8 months ago	Hiring Data(CSV)
		Sample Data_2 🛪 🖌 🖉 🗄	William Martin	8 months ago	8 months ago	Sample Data
		Hiring Data(CSV)_3	William Martin	8 months ago	8 months ago	Hiring Data(CSV)
		Hiring Data(CSV)_4	William Martin	8 months ago	8 months ago	Hiring Data(CSV)

- iii) The 'Import Preparation' window opens.
- iv) Select an option by marking the checkbox out of Local and VCS options.
- v) Browse a downloaded JSON file.
- vi) Select a dataset of similar metadata from the drop-down menu.
- vii) Click the '**Ok**' option.

Import Preparation	×
Local From VCS	
Hiring Data(CSV)_0.json	5 Browse
6 Select Dataset	_
Hiring Data(CSV)	-
;	Cancel Ok

- viii) A success message appears.
- ix) The Preparation gets imported and applied to the selected dataset.



T Import Preparation
Dataset
Hiring Data(CSV)
Hiring Data(CSV)
Hiring Data(CSV)
Hiring Data(CSV)
н

4.2. Datasets

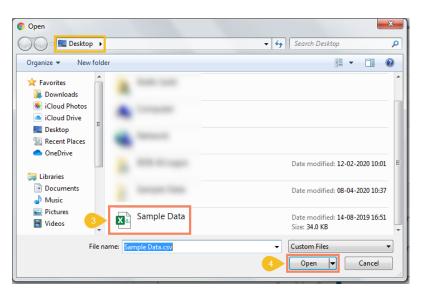
The '**Datasets**' section lists the data/inputs added to the system. The user can create a new preparation by selecting any of the listed datasets. The Datasets window also provides an option to add new datasets.

4.2.1. Adding a new Dataset

- i) Navigate to the Datasets option.
- ii) Click the 'Add Dataset' option.

Data Preparation						Take a Tour
🔂 Home						2 + Add Dataset
Preparations	Type here to search a dataset					م
	Name		Author	Created	Modified Date	Data Sampling Type
Datasets	New Microsoft Excel Workshee	✓ □ ± ■	William Martin	8 months ago	8 months ago	First 10K
	Sample Data	/ 🗈 ± 🔳	William Martin	8 months ago	8 months ago	First 10K
	Sample CSV File	∠ [] ± ∎	William Martin	8 months ago	8 months ago	Random
	Hiring Data(CSV)	∕ ⊡ ± ∎	William Martin	9 months ago	9 months ago	Random
	325419269	/ □ ± ∎	William Martin	10 months ago	10 months ago	First 10K
	492fbd404c42a043	Z 10 ± 1	William Martin	10 months ago	10 months ago	First 10K

- iii) A new window opens redirecting the user to select a CSV file.
- iv) Browse the file and upload it.





- v) The Data Set window appears with the selected CSV file.
- vi) The user can select a **Data Sampling Type** out of **'Random'** or **'First 10k'** options by marking the radio button.
- vii) The user can set an option to separate the data values in the selected file using the drop-down menu (E.g., Comma is selected in the following image).
- viii) Click the '**Ok**' option.

Data Set	×
5 Enter dataset name	
Sample Data	
6 Data Sampling Type: Random C First 10k	
56	
Note : Dataset if exists, will be overwritten.	Cancel Ok

- ix) A success message appears.
- x) The selected CSV File gets added to the Datasets page.

					Take a Tou
					+ Add Datas
Type here to search a dataset					
Name		Author	Created	Modified Date	Data Sampling Type
Sample Data	Ē 🛓 🚺	William Martin	a few seconds ago	a few seconds ago	Random
New Microsoft Excel Workshee	Ē 🛓 🔳	William Martin	8 months ago	8 months ago	First 10K
Sample Data	Ē 🖄 🔳	William Martin	8 months ago	8 months ago	First 10K
Sample CSV File	□ ± ■	William Martin	8 months ago	8 months ago	Random
Hiring Data(CSV)	© ± ∎	William Martin	9 months ago	9 months ago	Random
325419269	Dataset Added	Successfully!	ago	10 months ago	First 10K
	Name Sample Data New Microsoft Excel Workshee Sample Data Sample CSV File Hiring Data(CSV)	Name Sample Data New Microsoft Excel Workshee Sample Data CSV File Hiring Data(CSV)	Name Author Sample Data 	Name Author Created Sample Data Image: 1 million markin a few seconds ago New Microsoft Excel Workshee Image: 1 million markin 8 months ago Sample Data Image: 1 million markin 8 months ago Sample CSV File Image: 1 million markin 8 months ago Hiring Data(CSV) Image: 1 million markin 9 months ago	Name Author Created Modified Date Sample Data Image: Sample D

Note: The standalone version of the Data Preparation supports only CSV input of max 10k records. To work on other data sources and colossal volume, please use the ETL integrated version of data cleansing.

5. Data Grid

The Data Grid in the BDB Data Preparation is used for visualizing the data. The data displayed in the grid is a sample from the actual data set or complete data based on the data volume. The grid always shows the first 10 K rows in the dataset.

The user can access the Data Grid view of the selected dataset or data preparation by clicking on it. The displayed data in the grid changes based on the number of transforms performed on it.



a Pr	eparation						🌼 🗖	l 'D C'	▶ ▶ ←	-
							usd_billing			
	usd_billing = integer		source = string	experience_Year and integer	candidate_id integer	skills st	Profile	Transforms	Steps (0)	
1	1750	Male	Referral	3	148	Selenium	Chart	Info	Pa	atteri
2	3800	Male	Referral	9	150	SQL	Row Count			
3		Female	CareerNet	4	13	Selenium	60	, 		
4	2700	Male	Referral	5	28	Selenium				
5	1750	Male	CareerNet	2	17	Selenium	50	_		
6	2725	Female	BDB	5	217	BizViz, Manual QA				
7	1500	Male	CareerNet	2	112	Java	40 -			
8	2625	Male	BDB	4	200	Java				
9	2625	Male	BDB	4	207	Java+UI	30 -			
10	0	Male	Cosmic	4	161	SQL				
11	3700	Female	IvyPeople	11	151	Java	20 -			
12	2000	Female	Indeed	3	27	Selenium				
13	3000	Male	Referral	5	122	Analytics	10			

Note: The above image displays the data grid page opened when clicked on a dataset from the Datasets page.

5.1. Data Grid Header

The grid has a header that displays the column name and column type from the selected dataset.

	usd_billing inte	gender string		experience_Year = integer	candidate_id	skills
1	1750	Male	Referral	3	148	Selenium
2	3800	Male	Referral	9	150	SQL
3		Female	CareerNet	4	13	Selenium
4	2700	Male	Referral	5	28	Selenium
5	1750	Male	CareerNet	2	17	Selenium
6	2725	Female	BDB	5	217	BizViz, Manual QA
7	1500	Male	CareerNet	2	112	Java
8	2625	Male	BDB	4	200	Java
9	2625	Male	BDB	4	207	Java+UI
10	0	Male	Cosmic	4	161	SQL
11	3700	Female	IvyPeople	11	151	Java
12	2000	Female	Indeed	3	27	Selenium
13	3000	Male	Referral	5	122	Analytics
	1000					

Each Column Header has a Context menu \equiv icon. The Context menu displays the Column Type, option to Delete the column, and option to Rename.





It also presents the data type of the column. It is analyzed based on the max match to any data type in the first 10K records.

Consider that out of 10000 rows sample, there are 9000 integers and 1000 string values, the selected data type is Integer. The 1000 string rows get detected as invalid rows.

5.1.1. Data Types

The Data Grid header displays Data Types. It supports the following data types:

- 1. Integer
- 2. Double
- 3. String
- 4. Date
- 5. Timestamp

Data Preparation

	expected_joining =	previous_ctc =	team string	expyrsper_ctc
1	26-06-2017	0	BU 4	0
2	28-02-2017	300000	BU 7	300000
3	25-04-2017	0	BU 2	203125
4	22-05-2017	640000	BU 7	256410
5	30-10-2017	0	BU 7	233333
6	19-06-2017	1330000	BU 4	320000
7	07-06-2017	0	BU 7	425000
8	21-11-2017	407000	BU 4	285000
9	08-09-2017	0	BU 8	300000
10	12-06-2017	0	BU 4	212500
		(50000	8113	

5.2. Panel to List the Selected Filters.

The user can insert a filter condition by using the 'Search' option provided under the '**Chart**' tab. When a filter is selected, it gets added to the filter panel on the top of the data grid. The added filter has an option to remove it by clicking the '**Close**' (X) mark.

Steps to Apply Filter Condition

- i) Type the filter condition using the '**Search**' bar.
- ii) The Bar chart displays the searched value. Click on it to apply the filter.



	>I	gend	ler										
		Pro	ofile			Tra	nsfo	rms			Steps	s (1)	
		Ch	art				Info)			Patt	ern	
		Row (Count										
1	fen	nale										(Q
		Gro	up								Со	unt 🔺	
	0	5	10	15	20	25	30	35	40	45	50		1
2		Femal	e									46	

- iii) The applied filter condition displays on the top of the data grid table.
- iv) The selected column displays the filtered data.

		1					
gend	er= Female ×						
	usd_billing	≡ integer	gender	≡ string	source	≡ string	experience_Year integer
14	1000		Female		Drive		0
15	9		Female		Orgspire		5
16	2200		Female		CareerNet		3
20	2300		Female		Referral		5
22	1500	4	Female		BDB		1
23	1800		Female		CareerNet		3
25	3400		Female		Referral		6
26	2200		Female		IvyPeople		3
28	2300		Female		Referral		4
31	9		Female		CareerNet		3

v) The rows meeting the filter condition out of the total number of rows gets displayed on the bottomleft part of the page.

	74	2425	Female	BDB	4		218			
	78	1800	Female	BDB	2		201			
	81	4000	Female	Referral	12		160			
4(6/ 224]		« Pre	vious 1 2 3	4	5	12	Next	»

5.3. Data Quality Bar in the Grid

A Data Quality Bar appears in the header of the grid. The Data Quality is indicated through color-coding, as explained below:



• Dark Blue-Valid Data

skills=	Java, Big Data X	gender= Female ×			
	monthly_salary intege		name string	current_status string	designation string
78	40000	40000	Ranjana	Iransferred	Software Developer
97	40000	30000	Ishana	Transferred	Big Data Developer

• Orange– Invalid data

	offered_ctc	expected_joining ≡ date	previous_ctc	team = string	expyrsper_ctc = integer
1	0	26-06-2017	0	BU 4	0
2	300000	28-02-2017	300000	BU 7	300000
3	1300000	25-04-2017	0	BU 2	203125
4	1000000	22-05-2017	640000	BU 7	256410
5	700000	30-10-2017	0	BU 7	233333
6	1600000	19-06-2017	1330000	BU 4	320000
7	425000	07-06-2017	0	BU 7	425000
8	570000	21-11-2017	407000	BU 4	285000
9	300000	08-09-2017	0	BU 8	300000

• Light Blue -Blank data

Data Pr	eparation			6						
skills	= Java, Big Data	×	gender= Female	×						
	name	≡ string	current_status	≡ string	designation	string	referral_of	string	joining_status	≡ string
78	Ranjana		Transferred		Software Devel	oper			BizViz Core	
97	Ishana		Transferred		Big Data Devel	oper			BizViz Core	

5.4. Pagination

Pagination is implemented for the grid data. The tool displays 20 records on each page.

	1600000	19-06-2017	1330000	BU 4	320000
	425000	07-06-2017	0	BU 7	425000
8	570000	21-11-2017	407000	BU 4	285000
9	300000	08-09-2017	0	BU 8	300000
10	850000	12-06-2017	0	BU 4	212500
11	800000	27-03-2017	450000	BU 7	296296
12	300000	03-07-2017	204000	BU 4	200000
13	750000	16-08-2017	420000	BU 4	187500
14	300000	31-05-2017	0	BU 7	0
15	925000	12-05-2018	618000	BU 6	185000
16	550000	20-03-2017	350000	BU 7	196429
17	600000	03-07-2017	475000	BU 4	285714
18	900000	27-06-2017	0	BU 8	272727
19	1190400	01-12-2016	967200	BU 10	170057
20	650000	18-03-2018	580000	BU 6	130000

Note: The maximum rows displayed for sampling is always 10k.

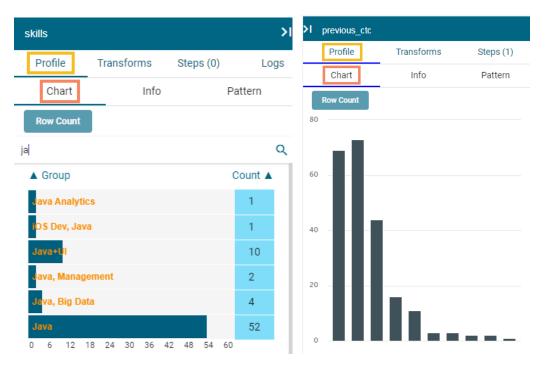


6. Summary Pane

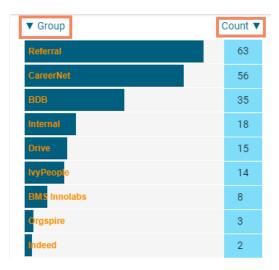
The summary pane gives an overview of the data profile like different patterns of data, distinct values, and occurrences.

6.1. Charts

The in-built charts (Column and Bar charts) display the occurrence of each value. The Bar appears to display string value. The Column chart projects numeric value columns and dates.



The chart can be sorted based on the group or the count of occurrence of a group. The sorted chart displays values from Ascending to Descending manner.



6.2. Info: Value/Statistics

The information tab displays the value or statistics of the data. The following aspects are displayed about the chosen data when the column is of string type:



- o Count: Count of Rows
- Valid: Count of Valid Data
- Invalid: Count of Invalid Data
- Duplicate: Count of Duplicates
- Distinct: Distinct Values

_	skills			
skills ≡ string	Profile	Transforms	Steps (1)	Log
Selenium	Chart	Info	Pat	tern
SQL	Count	219		
Selenium	Valid	219		
Selenium	Invalid	0		
BizViz, Manual QA	Duplicate:	189		
Java				
Java	Distinct:	30		
Java+UI				
SQL				
Java				
Selenium				
Analytics				
SQL				

When the selected column is of numeric type, other than the above-mentioned details the additional displayed information under the '**Info**' tab is based on aggregation functions as mentioned below:

- o Minimum
- o Maximum
- o Mean
- \circ Variance

	id			>
id ≡ integer	Profile	Transforms	Steps (1)	Logs
148	Chart	Info	Patte	ern
150	Count	219		
28	Valid	219		
17	Invalid	0		
217	MAX	224		
112				
200	MIN	1		
207	Mean	114.7	79	
161	Duplicate	0		
151	Distinct:	219		
27	Variance:	4,060	0.26	
122				
140				



6.3. Pattern

This section focuses on how data patterns and occurrences of each pattern in the dataset sample get plotted in a chart for the selected column.

		id						
id	≡ integer	Pr	ofile	Trans	forms	Step	s (1)	L
148			Chart	-	Info		Pa	attern
150								0.5
28		99						85
17		999)					125
217		9						9
112		0	40	80	120	160	200	
200								
207								
161								
151								
27								
122								
140								

Note: The value displayed is not the actual value, and it's just a pattern of the value.

6.4. Transforms

Data Preparation module provides a list of transforms that can be performed on the data to clean/prepare the data for insightful visualization.

This section explains the details of the transforms.

6.4.1. Advanced

6.4.1.1. Cluster & Edit

The '**Cluster & Edit**' transform when applied groups the words with similar phonetic (Speech sound/Pronunciation) into a cluster. The user can apply this transform to replace function on that bucket to replace all those words at once. We can also exclude some value when replacing it with the new value. It works on the Soundex algorithm to cluster the data.

When the Cluster & Edit transform gets applied as follows:

Cluste	r & Edit		- ×
Method Sounde:	х	 Soundex is a phonetic algorithm for indexing names by sound, as pronounced in English 	
	Values found	Replace Value	
	New_YO_Rk (1 rows)		
~	NEW YORKEE (1 rows)	New York	
	New York (1 rows)	TWO THE	
	Noo York (1 rows)		
_	Chicago (2 rows)		Ŧ
	CHICAGO (1 rows)	Chicago	Ť
	San Jose (1 rows)	San Jose	Ŧ
	San Whosay (1 rows)	San Whosay	T
	Boston (1 rows)		

The existing column 'Location' with the following Phonetic variations as displayed below:



Location			Location
New_York_City		1	New York
Bostn		2	Bostn
Chicago		3	Chicago
Nu Yorkk Sity		4	New York
New York		5	New York
Noo York		6	New York
NewYorkCity		7	New York
San Jose		8	San Jose
Chicago		9	Chicago
Boston		10	Boston
Nu Yorkk Sity,		11	New York
New_YO_Rk		12	New York
San Whosay		13	San Whos
NEW YORKEE		14	New York
Noo York City		15	New York
CHICAGO	it gets converted into	16	CHICAGO

6.4.1.2. Expression Editor

The Expression Editor transform has a collection of different functions to manipulate the data like absolute, to date, from Unix time.

- i) Select the 'Expression Editor' transform option using the 'Transforms' tab
- ii) The Expression Editor window opens with the following information:
 - a. Search Function- Use double click to search/select a function from the displayed list. The selected function appears under the '**Formula**' space.
 - b. Search Column- Use double click to search/select a column from the displayed list. The selected column appears under the '**Formula**' space.
 - c. The user can select an existing column by enabling the '**Update column**' option or create a New Column by entering the column name for the new column. (E.g., the following image displays New Column creation)
 - d. The selected function and column appear under the 'Formula' space.
 - e. Click the 'Submit' option.

Search function ABS ADD MONTHS AND BROUND CAST CBRT CONV CURRENT DATE CURRENT TIMESTA FROM UNIXTIME IF IN ISNULL LIKE	2 Search column usd_billing gender source experience_Vear candidate_id skills previous_organisation id offered_ctc expected_joining_date previous_ctc team exprysper_ctc monthly_salary	New Joining Date	
---	---	------------------	--

The new column gets added to the data grid with the updated data based on the applied formula.



joining_status == string	New Joining Date address
Joined	2020-04-13
BizViz Core	2020-04-13
Joined	2020-04-13
BizViz Core	2020-04-13
BizViz Core	2020-04-13
Declined	2020-04-13
Joined	2020-04-13

Note: In the case of selecting an existing column, the data gets updated as per the applied formula in the column.

E.g., the following image displays selecting an existing column.

Expression Editor			-
Search function	Search column	Q 🔽 Upd	date column
ABS	usd_billing	Select colum	
ADD MONTHS	gender		eu_johning_uate
AND	300.00	4 Formula current_	date()
BROUND	experience_Year		
CAST CBRT	candidate_id skills		
CONV	previous_organisation		
CURRENT DATE	id		
CURRENT TIMESTAMP	offered_ctc		
FROM UNIXTIME	expected_joining_date	I —	
IF	previous_ctc	Return	rns the current date
IN	team expyrsper_ctc		
LIKE	monthly_salary		
			5 Sub
			ocherate i innary reg
	expected_joining	g ≡ date	expected_joining 🧮 date
		dato	
	24-07-2017		2020-04-13
	20-09-2017		2020-04-13
	06-03-2017		2020-04-13
	20-05-2018		2020-04-13
	01-12-2016		2020-04-13
	03-07-2017		2020-04-13
	01-12-2016		2020-04-13
			2020-04-13
	01-12-2016		
	01-12-2016 14-08-2017		2020-04-13
			2020-04-13 2020-04-13
	14-08-2017		
	14-08-2017 27-07-2017		2020-04-13
he original da	14-08-2017 27-07-2017 17-04-2017 26-06-2017 07-08-2017		2020-04-13 2020-04-13

Copyright © 2015-2021 BDB



6.4.1.3. SQL Transform

This transform allows the user to write SQL Query against the table as we can write in any SQL editor. This transform requires the table name to be mentioned as '**InputDS**' in the query.

SQL Editor		
Search function	Q Search column Q Use dataset as tablename Select *,current_date() as current_date fro	m data aat
ABS	user_id	muataset
ADD MONTHS	date	
AND	date_time	
AS	items	
AVG	subject	
BROUND	company	
CAST	city_name	
CBRT	city	
CONV	order_no	
CURRENT DATE	total	
CURRENT TIMESTAMP	resaurant	
FROM UNIXTIME	delivery_charge	
F	packing_charge	
N	picklocation	
	and the second se	

Use syntax : "Select [*,function as new_column_name] from dataset " Example : Select *,abs(column1) as abs_value from dataset

It adds a new column along with the existing ones in the grid. When '*' is not specified in the query It just returns the current date.

region = string	current_date ≡ date
Bangalore	2019-08-16
Kochi	2019-08-16
Bhopal	2019-08-16
Hyderabad City	2019-08-16
Bangalore	2019-08-16
South Bengal	2019-08-16
Hyderabad City	2019-08-16
Kochi	2019-08-16
Chennai Region	2019-08-16
Calcutta	2019-08-16
Chennai Region	2019-08-16
Bangalore	2019-08-16
Hyderabad City	2019-08-16
Delhi	2019-08-16
Calcutta	2019-08-16
Chennai Region	2019-08-16



6.4.2. Anonymization

Anonymization is a type of information sanitization whose intent is privacy protection. It is a data processing technique that removes or modifies personally identifiable information.

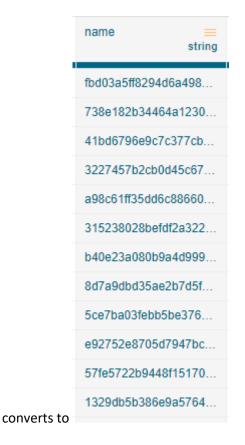
6.4.2.1. Data Hashing

Data Hashing is a technique of using an algorithm to map data of any size to a fixed length. Every hash value is unique. The supported Hash options are Hash, Sha-1, Sha-2 and MD-5.

Pro	Profile Transforms		Steps (9)	Logs
Co	olumn			
Type he	ere to se	arch a function		Q
Data H	^			
Hash Opti	ion:			
Hash				~
Hash				
Sha1				
Sha2				
MD5				
	aanningiii			

- a. Select the 'Data Hashing' transform from the 'Transforms' tab.
- b. Select a column from data grid for transformation
- c. Select the required 'Hash Option'- Hash, Sha-1, Sha-2, MD-5.
- d. Click the 'Submit' option.

The following image displays the transformed data after the 'Data Hashing' transform is applied.





6.4.2.2. Data Masking

Data masking transform is the process of hiding original data with modified content. It is a method of creating a structurally similar but inauthentic version of an actual **data**.

Select the 'Data Masking' transform from the 'Transforms' tab.

a. Select a column from data grid for transformation.

Column	
Type here to search a function	Q
Data Masking	^
Start Index:	
0	
End Index:	
3	\$
Subn	nit
Data Variance	~

- b. Select a start index and end index which you want to mask.
- c. Click the 'S**ubmi**t' option.

Below is the snapshot of how the 'Data Masking' transform when applied converts the selected data:

source	≡ string		source	≡ string
Indeed			****ed	
Orgspire			****pire	
Orgspire			****pire	
Referral			****rral	
Referral			****rral	
BMS Innolabs			****Innolabs	
Orgspire			****pire	
BMS Innolabs			****Innolabs	
Referral			****rral	
SkillRecruit			****IRecruit	
Orgspire			****pire	
Orgspire		gets converted to	****pire	



6.4.2.3. Data Variance

Select the 'Data Variance' transform from the 'Transforms' tab.

- a. Select a column from data grid for transformation.
- b. Select the required Value Type-Numeric/Date.
- c. Select an adequate operator from the options and a Percentage value.
- d. Click the 'Submit' option.

Column	
Type here to search a function	Q
Data Variance	^
Value Type:	
Numeric	~
Operator:	
+	~
Percentage:	
5	\$
Comments	
Add comments	
S	Submit

Below is the snapshot of how the 'Data Variance' transform when applied converts the selected data:

current_status	≡ string	experience	≡ integer
Transferred		15	
Resigned		10	
Terminated		4	
Transferred		5	
Transferred		2	
Declined		4	
Absconded		3	
Declined		3	
Declined		2	
Declined		2	
Absconded		5	
Declined		3	



6.4.3. Columns

6.4.3.1. Cast to Types

It is a table-based operation. The profiling of a column is done based on the data type present in the majority. Let's say in column A; we have four integer values and one string value, then the data type of column gets profiled as the integer despite one string value present in it. The 'Cast to Types' transform removes the value with the invalid data type. In this case, it converts data with a string data type to the null value.

Note: Cast to types is a lossy transformation. There is a possibility of some data loss.

6.4.3.2. Collect Set

The '**Collect Set**' transform generates the list of all the unique values of the column based on the selected column. It performs group concatenation.

Configure the Transform and click the 'Submit' option.

Column	
Find a Function	Q
Collect Set	
Create new column	
Partitioning Column	
Select Column *	
source	*
	Submit

It generates a list of all unique values as displayed in the below image:

team	≡ string	team_set_1	string
BU 8		[BU 8]	
BU 4		[BU 4]	
BU 7		[BU 7]	
BU 4		[BU 4]	
BU 4		[BU 4]	
BU 7		[BU 7]	
BU 6		[BU 6]	
BU 7		[BU 7]	
BU 4		[BU 4]	
BU 8		[BU 8]	
BU 10		[BU 10]	
BU 6		[BU 6]	



6.4.3.3. Concatenate with

The users can concatenate a column value with some other column or with some prefix/suffix. To perform the transform, select the column to which data must be concatenated and select the **'concatenate with'** transform. The available options are:

- a. Prefix: Specify the value to be prefixed to the selected column value
- b. Use with:
 - i. Select the 'Value' to add a Prefix/Suffix
 - ii. Select 'Other column' to concatenate two columns
- **c. Suffix:** Specify the value to be suffixed to the selected column value returns when performed on the selected column.

Concatenate with	
Create new column	
Prefix	
BDB_	
Use with	
Value	•
Suffox	
	Submit

The above configuration provides 'BDB_' as the prefix for the new column, 'Candidate_id_concat_1'.

	candidate_id ir	ateger candidate_id_conc ≡
1	105	BDB_105
2	192	BDB_192
3	62	BDB_62
4	70	BDB_70
5	170	BDB_170
6	92	BDB_92
7	121	BDB_121
8	169	BDB_169
9	182	BDB_182
10	102	BDB_102
11	21	BDB_21
12	112	BDB_112
13	97	BDB_97

The users must select 'Use with Other column' option to concatenate a value with another column and select the 'Use with Value' option to add prefix/suffix.

6.4.3.4. Delete Column

It deletes any selected column. To perform the transform, select the column and click on the '**Delete Column**' transform.

www.bdb.ai



6.4.3.5. Duplicate Columns

The 'Duplicate Columns' transform creates another column containing the duplicate data of the selected column.

skills 🗮 string		skills_duplicate_1 ≡ string
Java		Java
Java		Java
Scripting		Scripting
Java		Java
Java		Java
DotNet		DotNet
Java		Java
Java		Java
Selenium		Selenium
Java		Java
Java		Java
Java		Java
DotNet	returns	DotNet

6.4.3.6. Fill Empty

The '**Fill Empty**' transform is used to fill the null/empty value of cell using either above or below values available in the column.

Configure the 'Fill Empty' transform:

- i) **Create new column** Click the checkbox to create a new column or else the currently selected column gets updated.
- ii) Use with-The user can use either of the options from the provided choices:
 - a. From Above: To fill the empty cells and replace them by the value of the cells given above the empty cells.
 - b. From Below: To fill the empty cells and replace them by the value of the cells given below the empty cells.





items_replace = string		subject = string
		150,150,1 x,Chicken
		150,150,1 x,Chicken
		150,150,1 x,Chicken
150,150,1 x,Chicken		150,150,1 x,Chicken
		,Sambar Rice,1,66.66
,Sambar Rice,1,66.66	converts to	,Sambar Rice,1,66.66

6.4.3.7. Generate Primary Key

It generates the primary key for the table. It is a table-based operation. **Use with:** The user gets two options to generate the primary key:

- i) Contiguous- it generates the auto-incremented value starting from 1.
- ii) Non_contiguous- it generates a unique and random integer value.

Generate Primary Key	
Use with:	
Contiguous	▼
	Submit

A new column with primary values gets added to the data grid.

	Primary_column_1 ≡ integer
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13

6.4.3.8. Get Character Length

The transform '**Get Character Length**' when applied adds a new column with numbers displaying the length of character present in that cell.



designation string	designation_length_1 ≡ integer
QA Manager	10
QA Architect	12
Senior Software Engin	24
QA Engineer	11
QA Engineer	11
Senior Software Engin	24
AWS Consultant	14
Senior Software Engin	24
QA Engineer	11
Business Analyst	16
Senior QA Engineer	18
QA Engineer	11
Senior QA Engineer	18

The empty cells are kept as it is in the column.

items	≡ string		items_length_1	integer
			49	
1,Special Hyder	abadi			
			26	
150,150,1 x,Chi	cken			
,Sambar Rice,1,	66.66		20	
170,85,2,Masala	a Dosa	it returns	20	

6.4.3.9. Get JSON Objects

The 'Get JSON Objects' transform extracts any parameter from a given column with JSON data.

Sales_detials	
{"Category":"Beverages","	Product":"Chai", "Sales":"705.6", "Quarter"."Qtr 1"}
("Category":"Beverages","	Product": "Chai", "Sales": "878.4", "Quarter": "Qtr 2"}
{"Category":"Beverages","	Product":"Chai", "Sales":"1174.5", "Quarter":"Qtr 3"}
{"Category":"Beverages","	Product":"Chai","Sales":"2128.5","Quarter":"Qtr 4"}
{"Category":"Beverages","	Product":"Ipoh Coffee", "Sales":"1398.4", "Quarter": "Qtr 1"}
{"Category":"Beverages","	Product":"Ipoh Coffee", "Sales":"4496.5", "Quarter": "Qtr 2"}
{"Category":"Beverages","	Product":"Ipoh Coffee","Sales":"1196","Quarter":"Qtr 3"}
{"Category":"Beverages","	Product":"Ipoh Coffee", "Sales":"3979", "Quarter":"Qtr 4"}
{"Category":"Beverages","	Product":"Juices", "Sales":"1141.92", "Quarter":"Qtr 1"}
{"Category":"Beverages","	Product":"Juices","Sales":"1774.08","Quarter":"Qtr 2"}
{"Category":"Beverages","	Product":"Juices", "Sales":"3261.6", "Quarter":"Qtr 3"}
{"Category":"Beverages","	Product":"Juices", "Sales":"1705.5", "Quarter":"Qtr 4"}
{"Category":"Condiments",	"Product"."Aniseed Syrup","Sales"."544","Quarter"."Qtr 1"}
{"Category":"Condiments",	"Product":"Aniseed Syrup","Sales"."600","Quarter":"Qtr 2"}
{"Category":"Condiments",	"Product":"Aniseed Syrup","Sales":"140","Quarter":"Qtr 3"}
"Category":"Condiments",	"Product":"Aniseed Syrup", "Sales":"440", "Quarter":"Qtr 4"}



	Submit
Category,Product,Sales,Quarter	
Parameters to be extracted	
Get JSON Objects	^

When the parameter in the JSON is specified, the transform extracts all parameter values as columns into a tabular format.

Sales_detials	Sales_detials_Cat string	Sales_detials_Pro string	Sales_detials_Sales double	Sales_detials_Qua string
{"Category":"Beverages","Produ	Beverages	Chai	705.6	Qtr 1
{"Category":"Beverages","Produ	Beverages	Chai	878.4	Qtr 2
{"Category":"Beverages","Produ	Beverages	Chai	1174.5	Qtr 3
{"Category":"Beverages","Produ	Beverages	Chai	2128.5	Qtr 4
{"Category":"Beverages","Produ	Beverages	Ipoh Coffee	1398.4	Qtr 1
{"Category":"Beverages","Produ	Beverages	Ipoh Coffee	4496.5	Qtr 2
{"Category":"Beverages","Produ	Beverages	Ipoh Coffee	1196	Qtr 3
{"Category":"Beverages","Produ	Beverages	Ipoh Coffee	3979	Qtr 4
{"Category":"Beverages","Produ	Beverages	Juices	1141.92	Qtr 1
{"Category":"Beverages","Produ	Beverages	Juices	1774.08	Qtr 2
{"Category":"Beverages","Produ	Beverages	Juices	3261.6	Qtr 3
{"Category":"Beverages","Produ	Beverages	Juices	1705.5	Qtr 4
{"Category":"Condiments","Prod	Condiments	Aniseed Syrup	544	Qtr 1
{"Category": "Condiments", "Prod	Condiments	Aniseed Syrup	600	Qtr 2
{"Category": "Condiments", "Prod	Condiments	Aniseed Syrup	140	Qtr 3

To extract the nested Json the '.' mark can be used to specify the path.



The above data can be extracted using the transform by specifying the parameters as given below

Get JSON Objects	^
Parameters to be extracted	
name,age,cars.car1,cars.car2,cars.car3	
	Submit

The data is converted to a tabular structure as below

Vehicle_Info	Vehicle_Info_name string	Vehicle_Info_age integer	Vehicle_Info_cars string	Vehicle_Info_cars string	Vehicle_Info_cars string
{"name":"John","age":30,"	John	30	Ford	BMW	Fiat

6.4.3.10. Pivot

When applied, the 'Pivot' transform converts the data into a Pivot table based on the selected Pivot



Column and Group of the selected columns.

Sample Data

Category string	Product = string	Sales double	Quarter string
Condiments	Hot Pepper Sauce	1347.36	Qtr 1
Beverages	Ipoh Coffee	3979.00	Qtr 4
Beverages	Ipoh Coffee	1196.00	Qtr 3
Dairy Products	Yogurt	242.50	Qtr 2
Condiments	Grandma's Boysenberry Spread	1750.00	Qtr 3
Condiments	Hot Pepper Sauce	2150.77	Qtr 2
Condiments	Aniseed Syrup	440.00	Qtr 4
Confections	Chocolade	162.56	Qtr 2
Condiments	Northwoods Cranberry Sauce	1300.00	Qtr 2
Condiments	Hot Pepper Sauce	3857.41	Qtr 4
Beverages	Juices	1141.92	Qtr 1
Condiments	Aniseed Syrup	544.00	Qtr 1
Dairy Products	Yogurt	150.00	Qtr 4
Dairy Products	Mik	3329.28	Otr 1

In the given example, the 'Quarter' is selected as the Pivot Column, and Sum aggregation is selected for the Sales column. The Selected group columns are Category and Product.

Transform Selection:

Select Column	
Category, Product	-
Pivot Column:	
Quarter	•
Avg	Columns:
	Select Column 🔻
Count	Columns:
	Select Column 👻
Max	Columns:
	Select Column 🔹
Min	Columns:
	Select Column 👻
✔ Sum	Columns: Select Column
	Sales 💌

Thus, it returns the sum of Sales is aggregated based on Category and Product columns for each Quarter.



Category	tring	Product = string	Qtr 1 double	Qtr 2	Qtr 3 addressed	Qtr 4
Beverages		Chai	705.6	878.4	1174.5	2128.5
Beverages		Ipoh Coffee	1398.4	4496.5	1196.0	3979.0
Condiments		Aniseed Syrup	544.0	600.0	140.0	440.0
Dairy Products		Cheese	4454.8	174.15	2541.29	2472.5
Condiments		Northwoods Cranberry Sa		1300.0		2960.0
Dairy Products		Probiotics	487.0	2993.12	1458.75	2681.87
Condiments		Grandma's Boysenberry S			1750.0	750.0
Dairy Products		Yogurt	294.0	242.5	99.5	150.0
Confections		Chocolade	744.6	162.56	68.85	306.0
Dairy Products		Mik	3329.28	3989.9	10273.1	3060.0
Beverages		Juices	1141.92	1774.08	3261.6	1705.5
Condiments		Hot Pepper Sauce	1347.36	2150.77	1975.54	3857.41
Condiments		Vegie-spread	3202.87	263.4	842.88	2590.1
Condiments		Gula Malacca	1994.85	1753.62	1093.09	1701.87

6.4.3.11. Rename Column

The Rename Column transform allows the user to rename the selected column.

- a. Select a column from the data grid that needs to be renamed.
- b. Choose the 'Rename Column' transform from the Transforms tab.

	usd_billing = integer	gender ≡ string	source and string	experience_Year 🗮 integer	candidate_id anteger	Profile Transforms	Steps (0)
1	0	Male	IvyPeople	0	105	Column	
2	1500	Male	BDB	1	192	Find a Function	C
3	3400	Male	CareerNet	6	62	Concatenate with	
4	0	Male	Referral	4	70	Delete column	
5	2200	Male	IvyPeople	3	170	Duplicate Columns	
6	3000	Male	Referral	5	92		
7	1500	Male	Referral	1	121	Fill Empty	
8	1800	Male	Internal	2	169	Generate Primary Key	
9	1000	Male	Drive	0	182	Get Character Length	
10	2600	Male	CareerNet	4	102	Rename Column	
11	0	Male	BMS Innolabs	3	21	Return Non Null Column Values	
12	1500	Male	CareerNet	2	112	CONVERSIONS	
13	2600	Male	CareerNet	4	97	Convert Duration	

- c. The Rename Column dialog box opens. Provide a name that you wish to use as a rename for the selected column.
- d. Click the 'Submit' option.

Male	IvyPeople	0	105	
Rename Column	1			×
Rename this column				
				Submit
				Subilit

e. The column gets renamed.



	USD Billing	gender string
1	0	Male
2	1500	Male
3	3400	Male
4	0	Male
5	2200	Male
6	3000	Male
7	1500	Male
8	1800	Male
9	1000	Male
10	2600	Male
11	0	Male
12	1500	Male
13	2600	Male

6.4.3.12. Return Non-Null Column Values

The transform returns the first non-null value from the list of columns specified to a new column. To perform the transform, select the columns which must be checked for null and specify a column name for the result.

- a. Select Column: Select the columns to be checked for null
- b. Column name: The name for the new result column returns

Return Non Null Column Values	
Select Column *	
USD Billing, cur_monthly_payment	•
Column Name:	
Salary	
	Submit



	usd_billing 🗮 integer	cur_monthly_paym integer	
5	1750	43333	
6	1750		
7	2300		
8	2000		
9	2000		
10		52000	
11		52000	
12		52000	
13		43333	
14	4000	141666	
15		0	
16	1800	70833	
17	1750	38333	

	Salary integer
	1750
	1750
	2300
	2000
	2000
	52000
	52000
	52000
	43333
	4000
	0
	1800
returns the new result column	1750

6.4.4. Conversions

6.4.4.1. Convert Duration

The transform converts any duration (day, hour, minute, seconds, milliseconds) to any specified duration.

To perform the transform, select the column which has the duration to be converted and specify the duration type.

- a. From: The type of source interval
- b. **To:** The type of destination interval
- c. **Precision:** The decimal points to be retained
- d. Click the 'Submit' option.

Convert Duration	
Create new colu	mn
From	
Hour	•
То	
Minute	•
Precision	
1	
	Submit

Below is the snapshot of how the 'Convert Duration' transform when applied converts the selected data:



Duration_hrs double		Duration_hrs double
2.8		168.0
3.6		216.0
5.4		324.0
6.2		372.0
7.4		444.0
9.1		546.0
4.4		264.0
6.7		402.0
8.1		486.0
4.5		270.0
9.2	converts to	552.0

6.4.5. Data Cleansing

6.4.5.1. Clear Cells on Matching Value

Clear the cell value on matching the condition specified. Operators include contains, equals, starts with, end with, and regex match. Transform applies in the same column.

- **Operator:** Select the operator required for matching from the list
- Value: The value or pattern to be searched for in the selected column

Clear cells on matching value	
Operator:	
Equals =	•
Value:	
1	
-	
	Submit

The value selected in the form clears the cell with 1 in the selected column.



Gender ≡ integer		Gender string
male		male
female		female
male		male
0		0
1		
1		
1		
female		female
0		0
1		
male	returns data like this	male

6.4.5.2. Delete Rows on Matching Value

Delete the rows on matching the condition specified for that column. Operators include contains, equals, starts with, ends with, and regex match.

- **Operator:** Select the operator required for matching from the list
- Value: The value or pattern to be searched for in the selected column

Delete rows on matching value	
Operator:	
Regex */	•
Value:	
[0-9]	
	Submit

The value selected in the form deletes the row with any numbers from 0-9 in the selected column.

Gender 🗮 integer			
male	-		
female			
male			
0			
1		Gender	=
1			string
1		male	
female		female	
0		male	
1		female	
male	turns to	male	

when the above transform is applied.



6.4.5.3. Delete Rows with Empty Cell

a. The transform deletes any row which has a blank value in the selected column. The transform does not have a form.

Team	string	Designation = string	Referral_of string
B3		QA Lead	
B4		Software Eng.	
B6		Sr. Sofware Eng.	EMP 9
B8		Sr. Software Eng.	
B4		QA Eng.	
B1		Project Manager1	
B4		Executive Manager	EMP5
B5		BI Lead	
B6		Sr. Sofware Eng.	
B1		Project Manager2	
B8		Sr. Sofware Eng.	

b. When we perform the transform on column "referral_of" it deletes all the rows which have an empty value in that column returning the data as below:

Team	≡ string	Designation	≡ string	Referral_of	≡ string
B 6		Sr. Sofware Eng.		EMP 9	
B4		Executive Manage	er	EMP5	

6.4.5.4. Delete Rows with Invalid Cell

- a. The transform deletes any row which has an invalid value in the selected column. The transform does not have form.
- b. When we do the transform on the 'gender' column, it deletes all rows marked invalid as displayed below:



Gender	≡ string
male	
female	
male	
female	
1	
1	
female	
0	
1	
male	

6.4.5.5. Delete Rows with Negative Values

- 1. It deletes the rows which have a negative value in the selected column. This transform does not have a form.
- 2. When this transform is applied to experience column, it deletes all rows with negative, as displayed below:

Referral_of	≡ string	Experience	integer
		4	
		6	
EMP 9		6	
		7	
		-1	
		8	
EMP5		5	
		5	
		6	
		9	
		6	

3. It deletes the row with the negative value and returns the data as displayed below:



Referral_of	string	Experience	integer
		4	
		6	
EMP 9		6	
		7	
		8	
EMP5		5	
		5	
		6	
		9	
		6	

6.4.5.6. Fill Cells with Value

It fills the selected column with a value or a value from another column.

Type here to search a function	Q
Fill cells with value	^
Use with:	
Other column	•
Column:	
expected_joining_date	•
	Submit

- Use with: Specify whether to fill with a value or another column value
- **Column/ Value:** The value with which the column must be filled, or the column with which the value must be replaced

When the above transform is applied to the below data on the column 'created_date,' it copies the value from the 'expected_joining_date' column to the 'created_date' column.

expected_joining ≡ date	created_date	tring
24-07-2017		
03-07-2017		
14-08-2017		
08-09-2017		
15-06-2017		
21-08-2017		
10-07-2017		
04-05-2017		
10-04-2017		
01-12-2016		
19-06-2017		
01-12-2016		
20-11-2017		



Note: The user can also fill the column with a specific value if the selected option is 'Value'. E.g., the following image mentions 30 as the selected value for the 'created_date' column.

			created_date	≡ integer
			30	
			30	
Fill cells with value			30	
			30	
			30	
Fill collo with volue	^		30	
Fill Cells With Value			30	
Use with:			30	
Value	•		30	
Value:			30	
30			30	
	Submit		30	
		the column displays	30	

6.4.5.7. Fill Empty Cells with Text

It helps to fill the empty cells of a selected column with a value or a value from another column if the destination column is empty.

Fill empty cells with text	^
Use with:	
Value	•
Value:	
NA	
	Submit

- Use with: Specify whether to fill with a value or another column value.
- **Column/ Value:** The value with which the column must be filled, or the column with which the value must be replaced.

When the transform is applied to the below data on column 'referral_of,' it fills the value '**NA**' for all the empty cells of that column.

gnation ≡ string	referral_of string		designation 🗮 string	referral_of
e Engineer		1	Software Engineer	NA
ware Engineer			Software Engineer	NA
l Software Engineer			Lead Software Engineer	NA
nior Software Engin	Mahendra		Senior Software Engin	Mahendra
nior Software Engin			Senior Software Engin	NA
nior Software Engin	Manish Jaiswal		Senior Software Engin	Manish Jaiswa
sociate Software En	Ritesh		Associate Software En	Ritesh
ftware Engineer	Tripura		Software Engineer	Tripura
sociate QA Engineer			Associate QA Engineer	NA
enior Software Engin			Senior Software Engin	NA
enior Software Engin			Senior Software Engin	NA
ssociate Software En			Associate Software En	NA
enior Software Engin		converts to	Senior Software Engin	NA



6.4.5.8. Find Anomaly

Anomaly detection is used to identify any anomaly present in the data. i.e., Outlier. Instead of looking for usual points in the data, it looks for any anomaly. It uses the **Isolation Forest** algorithm.

The 'Find Anomaly' transform takes four parameters:

- 1. Select Feature Columns: We can select one or more columns where we want to find the anomaly.
- **2. Maximum Sample Size:** Isolation forest takes the training data of a given sample size to find out the normal value in the dataset. The sample size can vary from 1 to 250 (both inclusive).
- **3.** Contamination (%): It is the percentage of observations we believe to be outliers. It varies from 0 to 1 (both inclusive).
- 4. **Anomaly Flag Name:** The result is either 0 or 1. 0 means the data is standard, and 1 means data is an outlier. This information gets stored in the new column given in the anomaly flag name.
- 5. Click the 'Submit' option to detect anomaly from the selected data.

Find Anomaly	^
Select Feature Columns *	
value	*
Maximum Samples Size:	Contamination %
3	0.5
Anomaly Flag Name:	
Outliers	
	Submit

The anomaly gets stored in the new column under the anomaly flag name (In this case, it is displayed under the '**outlier**' column).

value	≡ integer	outlier	≡ double
1		0.0	
2		0.0	
3		0.0	
4		0.0	
21		1.0	
6		0.0	
1000		1.0	
1200		1.0	
1000		1.0	
		1.0	
		1.0	
		1.0	
		1.0	



6.4.5.9. Flag Duplicates in Columns

This transform adds a new Boolean column based on duplicate values in the column. For original value it gives false, and for the duplicate value, it provides true value.

	team string		IsDuplicate_team boolean
	BU 4		false
	BU 7		false
	BU 2		false
	BU 7		true
	BU 7		true
	BU 4		true
Flag Duplicates In Columns	BU 7		true
	BU 4		true
Select Column *	BU 8		false
team 👻	BU 4		true
Cubit	BU 7		true
Submit	BU 4	returns	true

6.4.5.10. Flag Duplicates in Tables

This transform adds a new Boolean column based on duplicate rows in the table. For original value it gives false, and for the duplicate value, it provides true value.

6.4.5.11. Remove Duplicates from Column

It removes duplicate values from the selected columns. This transform can be performed on a single as well as on multiple columns.

	team string		team str	≡ ing
	BU 4		BU 4	
	BU 7		BU 7	
	BU 2		BU 2	
Remove Duplicates From Column	BU 7		BU 8	
Remove Dupileates From Column	BU 7		BU 6	
Select Column *	BU 4		BU 10	
team 💌	BU 7		BU 5	
	BU 4		BU 1	
Submit	BU 8	converts to	BU 9	

6.4.5.12. Remove Duplicates from Table

It removes all duplicate rows from the table.



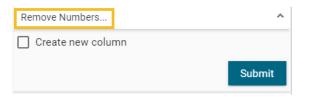
6.4.5.13. Remove Letters

It removes any letter present in the selected column. The users can either add a new column with the transformed value or overwrite the same column.

Type here to search a function.		Q			
Remove Letters		^			
Create new column					
	Su	ıbmit			
	Employee ID		E 1 10	_	
	string		Employee ID	≡ string	
	EMP ID 1		1		
	EMP ID 2		2		
	EMP ID 3		3		
	EMP ID 4		4		
	EMP ID 5		5		
	EMP ID 6		6		
	EMP ID 7		7		
	EMP ID 8		8		
	EMP ID 9		9		
	EMP ID 10		10		
The selected column	EMP ID 11	converts into	11		after transforma

6.4.5.14. Remove Numbers

It removes any number present in the selected column. We can either add a new column with the transformed value or overwrite the same column.



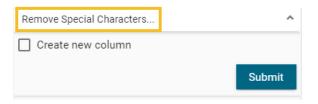
When the 'Remove Numbers', transform gets performed on a selected column,

Qualification string		Qualification
BE 1		BE
BE 2		BE
BE 3		BE
BE 4		BE
BE 5		BE
ME 6		ME
MTech 7		MTech
BTech 8		BTech
BE 9		BE
BTech 10		BTech
MTech 11	it removes numbers from the selected column	MTech



6.4.5.15. Remove Special Characters

It removes any special character present in the selected column. Only letters, numbers, and spaces are retained. We can either add a new column with the transformed value or overwrite the same column.



When the transform 'Remove Special Characters' gets performed on the selected column,

omments 📃 string		Comments
ot happy with the offfer.		not happy with
ot happy with the offer.		not happy with
Not Happy with the offer,		Not Happy with
Accepted it.		Accepted it
l am fine with it.		I am fine with it
I am fine with the offer.	the punctuations get removed from the column	I am fine with th

6.4.6. Dates

6.4.6.1. Add Duration

The transform adds two-time values. It can either add the selected column with a time value or time from another column. The transform supports adding time into 'hh:mm:ss.mmm' and 'hh:mm:ss' formats.

- Use with: Specify whether to fill with a value or another column value
- **Column/ Value:** The value with which the column must be added, or the column with which the selected column value must be added.

Add Duration	
Create new column	
Use with:	
Other column	•
Column:	
Shot1_duration	•
	Submit

The transform when performed on the data selecting 'Shot1_duration', it adds



Shot1 duration and Shot2	_duration and gives a new column with the result.

ihot 1 string			Shot 1 = string	Shot 2 = string	Shot 2_
	00:00.0		00:00.0	00:00.0	00:00:0
0.0	00:00.0		00:00.0	00:00.0	00:00:0
0.0	00:00.0		00:00.0	00:00.0	00:00:0
D1.0	00:01.0		00:01.0	00:01.0	00:02:0
:02.0	00:00.0		00:02.0	00:00.0	00:02:0
3.1	00:00.0		00:03.1	00:00.0	00:03:0
00.0	00:00.0		00:00.0	00:00.0	00:00:0
:00.0	00:02.0		00:00.0	00:02.0	00:02:0
00:01.0	00:02.0		00:01.0	00:02.0	00:03:0
00:02.1	00:00.0	converts to	00:02.1	00:00.0	00:02:0

6.4.6.2. Add Interval to Date

It adds the time duration specified to the selected datetime column.

- Input Format: It is used to specify the format of the selected Date column format. It can have values 'Year first', 'Month first', and 'Day first.'
- Value Type: It specifies the type of duration which acts as the operand for the addition. The value type can be years, months, days, weeks, hours, minutes or milliseconds
- Value: The value or the operand that must be added with the selected column

Note: The transform supports the datetime column of 'yyyy-mm-dd' into the 'hh:mm:ss' format.

6.4.6.3. Extract Time

Extract the time units from a selected column with a time value. The time units that get extracted include hours, minutes, seconds, milliseconds, and time to milliseconds.

- Hours: Extracts hours from a time
- Minutes: Extracts minutes from a time
- Seconds: Extracts seconds from a time
- MilliSeconds: Extracts milliseconds from a time
- Time to MilliSeconds: Converts the time given to milliseconds

Note : The transform supports time format like- hh:mm:ss:mmm, hh:mm:ss, hh:mm

6.4.6.4. Extract Date

It extracts the date part from a selected column with a date value. The date parts that can be extracted include day, month, year, the day of the week, the day of the year and the week of the year.

- Day: It extracts day from a date
- **Month:** It extracts the month from a date/datetime. We can specify the pattern in which the month value has to be returned. Month pattern can be 0-12, Jan Dec or January December
- **Year**: It extracts the year from a date. We can specify the pattern in which the year has to be returned. The year pattern can be in the 'yy' or 'yyyy' format.



- **Day of Week:** It returns the day of the week for the selected date. Day of week pattern can also be specified. The pattern can be 1-7, Sun-Sat or Sunday-Saturday
- **Day of Year:** It returns a number between 1 and 365, which indicates the sequential day number starting with day one on January 1st.
- Week of Year: It replaces a number between 1 and 53, which indicates the sequential week number beginning with 1 for the week January 1st falls.

Note: The transform supports Date and DateTime format (date hh:mm:ss)

6.4.6.5. Find Date Difference

The transform finds the difference between two date values. It can either subtract the selected column with a date value or date from another column. The transformed value can replace the existing column value or can be added as a new column.

- Input Format: Specifies the format of the given date column
- Use with: Specify whether to fill with a value or another column value
- Value Hint: Specifies format of value from which we want to find the difference
- Value: Pass the date value from where you want to find the date difference

Find Date Difference	^
Create new column	
Input Format:	
Year First	•
Use with:	
Value	•
Value Hint:	
Year First	•
Value:	
2016-01-01	
	Submit

This transform gives the number of days by finding out the difference between the given date and value/date column which we have used. Here value used is: 2016-01-01

expected_joining date		expected_joining ≡ integer
2017-05-22		507
2017-06-19		535
2017-07-06		552
2017-11-21		690
2017-06-27		543
2018-03-18		807
2017-06-03		519
2017-10-08		646
2017-06-26		542
2017-09-10		618
2017-06-26		542
2017-08-30		607
2017-09-08	converts to	616



6.4.6.6. Format Date

The users can change the format of a date column by using this transform.

- Source Format Hint: Specifies the current format of the date column.
- Target Format: Specifies what we want first(Year, Month, Day) in our output format of the date column
- Year Pattern: Specifies the format of the year (yyyy or yy) in the output date column.
- Month Pattern: It specifies the format of the month (number, Jan-Dec, January-December) in the output date column.
- **Delimiter:** Specifies Delimiter(like- slash, a hyphen, comma, full stop, space) for the output date column.
- Include Timestamp: It adds a timestamp to the current date format if enabled with a tick mark.

Format Date		^
Source Format Hint:	Target Fo	ormat:
Year First	▼ Year F	irst 🔻
Year Pattern:	Month Pa	attern:
уууу	▼ Jan-De	ec •
Delimiter:		
/		•
🔽 Include Time	stamp	
_		Submit
expected_joining date		expected_joining ≡ timestamp
2017-05-22		2017/May/22 00:00:00
2017-06-19		2017/Jun/19 00:00:00
2017-07-06		2017/Jul/06 00:00:00
2017-11-21		2017/Nov/21 00:00:00
2017-06-27		2017/Jun/27 00:00:00
2018-03-18		
		2018/Mar/18 00:00:00
2017-06-03		2018/Mar/18 00:00:00 2017/Jun/03 00:00:00
2017-06-03 2017-10-08		
		2017/Jun/03 00:00:00
2017-10-08		2017/Jun/03 00:00:00 2017/Oct/08 00:00:00
2017-10-08 2017-06-26		2017/Jun/03 00:00:00 2017/Oct/08 00:00:00 2017/Jun/26 00:00:00
2017-10-08 2017-06-26 2017-09-10		2017/Jun/03 00:00:00 2017/Oct/08 00:00:00 2017/Jun/26 00:00:00 2017/Sep/10 00:00:00

6.4.6.7. From Unix Time

The '**From Unix Time**' transform converts the Unix time into a specified format (1349862300 – 2012 10-10 both date and datetime).



6.4.6.8. Sub Interval to Date

The '**Sub Interval to Date**' transform subtracts specified value(interval) from the given date column. The transformed value can replace the existing column value or can be added as a new column.

- Input Format- Format of date column(given) should be specified here.
- Value Type-specifies what we want to subtract like years, months, days, weeks, etc.
- Value- specifies how many years/months/days(value type) we want to subtract.

Sub Interval To Date	^
Create new column	
Input Format:	
Year First	•
Value Type:	
Days	•
Value:	
10	
	Submit

This transform when performed subtracts four months from the date column and gives this new column having the date which is 10 days back from the given date.

expected_joining date		expected_joining ≡ date
2017-05-22		2017-05-12
2017-06-19		2017-06-09
2017-07-06		2017-06-26
2017-11-21		2017-11-11
2017-06-27		2017-06-17
2018-03-18		2018-03-08
2017-06-03		2017-05-24
2017-10-08		2017-09-28
2017-06-26		2017-06-16
2017-09-10		2017-08-31
2017-06-26		2017-06-16
2017-08-30		2017-08-20
2017-09-08	converts to	2017-08-29

6.4.6.9. Subtract Duration

The transform 'Subtract Duration' deducts the time values in two ways. It can either subtract the selected column with a time value or time from another column. The transform supports subtracting time into '**hh:mm:ss.mmm'**, '**hh:mm:ss'** and '**hh:mm'** formats. The transformed value can replace the existing column value or can be added as a new column.

- Use with: Specify whether to fill with a value or another column value
- **Column/ Value:** The value with which the column must be subtracted, or the column with which the selected column value must be subtracted.



Subtract Duration	^
Create new colu	mn
Use with:	
Other column	•
Column:	
Time_Split	•
	Submit

This transform when performed on Time1_split1 for subtracting 01:00:00 from this column provides a new column having values after deducting 01:00:00.

Time_Split	string	Time_Split ≡ string
01:00:00		00:00:00.000
02:00:00		00:00:00.000
03:00:00		00:00:00.000
04:00:00		00:00:00.000
05:00:00		00:00:00.000
06:00:00		00:00:00.000
07:00:00		00:00:00.000
08:00:00		00:00:00.000
09:00:00		00:00:00.000
10:00:00		00:00:00.000
11:00:00	converts	to 00:00:00.000

6.4.7. Integer

6.4.7.1. Add, Multiply, Subtract or Divide

It performs the arithmetic operation on the selected numerical column.

- **Operator:** There are four arithmetic operations to choose (+, -, /, *).
- Use with: The operation can be performed between column-column and column-value.
- **Operand/Column:** The arithmetic operation needs two operands. The first operand is one on which the operation is being performed. The second operation can either be a value or other numerical column based on the choice of use with an option.

Add,multiply,subtract or divide	^
Create new column	
Operator	
x	•
Use with:	
Value	•
Operand	
1000	\$
	Submit



Price(K) ≡ integer		Price(K)_multiply_1 ≡ integer
34		34000
176		176000
324		324000
74		74000
109		109000
111	converts to	111000

6.4.8. ML

6.4.8.1. Binarizer

It converts the value of a numerical column to zero when the value in the column is less than or equals to the threshold value and one if the value in the column is greater than threshold value.

	Screen Size adduble		Screen Size_binari = double
	13.3		0.0
	13.3		0.0
	15.6		1.0
	15.4		1.0
Binarizer	13.3		0.0
Threshold:	15.6		1.0
13.3	15.4		1.0
Submit	13.3	converts to	0.0

6.4.9. Numbers

6.4.9.1. Max

It gives the maximum value from the selected columns row-wise. The selected column should be numerical and more than one.

6.4.9.2. Mean

It gives the average value of the selected columns row-wise. The selected column should be numerical and more than one.

6.4.9.3. Min

It gives the minimum value from the selected columns row-wise. The selected column should be numerical and more than one.



6.4.9.4. Negate

It complements the sing of a numeric value. If the value is positive, then a negative value comes and vice-versa.

6.4.9.5. Number Name

It converts the value of the selected column into words. The column must be of integer type. **Use with:** It gives the users an option to convert word into either western format or Indian format.

	Price Euros 🔤 integer		Price Euros_In_W
Number Names	34900		Thirty Four Thousand Nine Hundred
✓ Create new column	176900		One Hundred and Seventy Six Thousand Nine Hundred
Use with:	324000		Three Hundred and Twenty Four Thousand
Western •	74900		Seventy Four Thousand Nine Hundred
Submit	109900	converts to	One Hundred and Nine Thousand Nine Hundred

6.4.9.6. Remove Fractional Part

It removes the fractional part from the numerical column. The float column is converted into the integer data type.

6.4.9.7. Round Value using Ceil Mode

It replaces the number with a greater integer value if the number is between two integer values. The transformed value can replace the existing column value or can be added as a new column.



6.4.9.8. Round Value using Down Mode

It rounds the number down to a specified digit or gives the specified number of decimals without any change in value. The transformed value can replace the existing column value or can be added as a new column.



	Suicide_per_100k doub		Suicide_per_100k ≡ integer
Pound value using down mode	6.71		6
Round value using down mode	5.19		5
Create new column	4.83		4
Precision:	4.59		4
0	3.28		3
Submit	2.81	converts to	2

6.4.9.9. Round Value using Floor Mode

It replaces a number with the lesser integer value, if the number is between two integer value, or it rounds the number down to the nearest multiple of Specified significance. It does not consider whether the next digit is 5 or less than or greater than 5. The transformed value can replace the existing column value or can be added as a new column.

		Suicide_per_100k double		Suicide_per_100k ≡ double
	~	6.71		6.7
Round value using floor mode		5.19		5.1
✓ Create new column		4.83		4.8
Precision:		4.59		4.5
1		3.28		3.2
	Submit	2.81	converts to	2.8

6.4.9.10. Round Value using Half-up mode

It replaces a number with the next integer value if its next digit is 5 or greater than 5. The transformed value can replace the existing column value or can be added as a new column.

	Suicide_per_100k double		Suicide_per_100k ≡ double
	6.71		6.7
Round value using halfup mode	5.19		5.2
Create new column	4.83		4.8
Precision:	4.59		4.6
1	3.28		3.3
Submit	2.81	converts to	2.8

6.4.10. String

6.4.10.1. Change to lower case

It converts the selected column value to the small case. The transformed value can replace the existing column value or can be added as a new column.



6.4.10.2. Change to Title Case

It converts the selected column value to the title case. The transformed value can replace the existing column value or can be added as a new column.

6.4.10.3. Change to Upper Case

It converts the selected column value to capital letters. The transformed value can replace the existing column value or can be added as a new column.

6.4.10.4. Extract Substring at Position

It extracts the substring from the selected column based on the starting position and the length of the extract. The transformed value can replace the existing column value or can be added as a new column.

- **Position:** This value is required and is the start position. It can be both a positive or negative number. If it is a positive number, this function extracts from the beginning of the string. If it is a negative number, this function extracts from the end of the string.
- Length: This value is optional. It specifies the number of characters to extract. If omitted, the whole string is returned starting from the given position.

6.4.10.5. Extract Substring before Delimiter

It extracts the substring from the selected column, before the 'n^{th'} occurrence of the delimiter specified where 'n' is the count. The transformed value can replace the existing column value or can be added as a new column.

- Delimiter: The delimiter on whose occurrence the extract should happen.
- **Count:** This value is mandatory and specifies the count of occurrence of the delimiter before which the extract should happen.

6.4.10.6. Insert Character

It inserts the character entered after a specified position. The transformed value can replace the existing column value or can be added as a new column.

- **Position:** The position in the cell value, after which the character must be inserted. We can even pass comma separated values. E.g., 2,4,6 insert the specified character after position 2, 4 & 6 of the cell values
- Character: The character that should be inserted after the specified positions

Insert Character	
Create new column	
Position:	
Character:	
	Submit



6.4.10.7. Remove Consecutive Characters

The transform removes the repeated whitespace or character and modifies the selected column /adds the result to a new column. It removes only the repetition.

- **Separator**: it has values whitespace /other. If whitespace, the transform searches for multiple white spaces and returns a single-spaced value.
- **Custom repeated Character:** When a repeated character is '**Other**,' this provides an option to give the character whose consecutive occurrence must be searched.

6.4.10.8. Remove Part of Text

It matches and removes the matching part or entire value based on the condition. The transformed value can replace the existing column value or can be added as a new column.

- Operator: Select the operator required for matching from the list
- Value: The value or pattern to be searched for in the selected column

6.4.10.9. Remove Trailing and Leading Characters

It removes trailing and leading characters from the column. The transformed value can replace the existing column value or can be added as a new column.

- **Padding character:** Specify whether to remove whitespace or another character using the dropdown menu.
- **Custom padding character -** If 'other' is selected as a padding character, specify which is the character to be removed.

Remove trailing and leading characters	
Create new column	
Padding character	
Other	•
Custom padding character:	
	Submit

6.4.10.10. Search and Replace

It searches and replaces the matching part or entire value based on the option selected. The transformed value can replace the existing column value or can be added as a new column.

Operator- Select the operator required for matching from the list. Operators include contains, equals, starts with, end with, and regex match.

Value: It is the value or pattern to be searched for in the selected column.



Search and replace	
Create new column	
Operator: Regex ^/ Search for:	•
Replace with:	
Overwrite entire cell	Submit

6.4.10.11. Split String

It splits the string based on condition. It displays new columns based on the number of delimiter and on position.

- Use With: Specify whether to split with a delimiter or at position
- **Delimiter:** The delimiter on whose occurrence the split should happen
- Position: After which position split should happen if use with is 'position.'

Split String	
Use with: Delimiter	•
Separator:	
	Submit

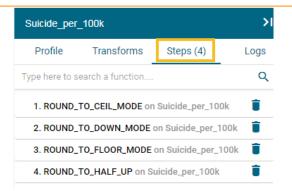
Here splitting of the column is done based on position (after the 5th character)

age ≡ strin		age 🚃 string	age_split_1 = integer	age_split_2 str
15-24 years		15-24 years	15	24 years
35-54 years		35-54 years	35	54 years
15-24 years		15-24 years	15	24 years
75+ years		75+ years	75+ years	
25-34 years		25-34 years	25	34 years
75+ years		75+ years	75+ years	
35-54 years	converts to	35-54 years	35	54 years

6.5. Steps

This tab lists all the transforms that were performed on the data. It also gives a count of steps performed.





The user can open any performed transform and edit it using the 'Steps' tab.

Suicide_per_	_100k		>	
Profile	Transforms	Steps (4)	Logs	
Type here to s	earch a function		Q	
1. ROUND_	TO_CEIL_MODE on	Suicide_per_100k		
2. ROUND_TO_DOWN_MODE on Suicide_per_100k 🏮				
3. ROUND_TO_FLOOR_MODE on Suicide_per_100k				
4. ROUND_	TO_HALF_UP on Su	iicide_per_100k	Î	
🔽 Create r	new column			
Precision:				
1				
		Su	ıbmit	

6.5.1. Deleting or Undeleting Steps

The 'Steps' tab has option to delete and undelete options to delete or undelete performed steps.

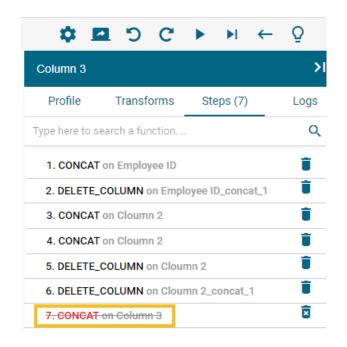
The user can follow the below given steps to use these options:

- i) Navigate to the 'Steps' tab (Make sure there are some steps listed on the tab).
- ii) Click the 'Delete' icon for a listed steps.

→ I4 < D C 🖪 🌣	Q
Column 3	N
Profile Transforms Steps (7)	Logs
Type here to search a function	Q
1. CONCAT on Employee ID	Î
2. DELETE_COLUMN on Employee ID_concat_1	Î
3. CONCAT on Cloumn 2	Î
4. CONCAT on Cloumn 2	Î
5. DELETE_COLUMN on Cloumn 2	Î
6. DELETE_COLUMN on Cloumn 2_concat_1	Î
7. CONCAT on Column 3	0



iii) The selected steps get deleted.



iv) A message appears to notify the same.

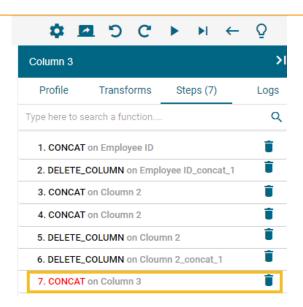
Step Deleted Successfully and Exclude from Execution

v) Click the 'Undelete' icon for the deleted step.

H - + D C 🖪 🕸	Q →
Column 3	>
Profile Transforms Steps (7)) Logs
Type here to search a function	Q
1. CONCAT on Employee ID	Î
2. DELETE_COLUMN on Employee ID_cond	at_1 🔋
3. CONCAT on Cloumn 2	Î
4. CONCAT on Cloumn 2	Î
5. DELETE_COLUMN on Cloumn 2	Î
6. DELETE_COLUMN on Cloumn 2_concat	.1
7. CONCAT on Column 3	

vi) The deleted step gets undeleted.





vii) A message appears to notify the same.



7. Navigation Pane

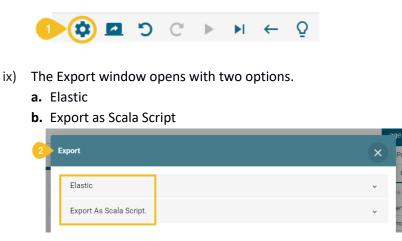
The navigation pane provides options to export the preparation steps in Elastic settings, move the steps out of the BDB Data Preparation. The navigation panel also has icons to perform Undo, Re-do, Replay Dirty, and Replay All options.



7.1. Export Steps to a Data Store Meta Data

The 'Export Steps to data store' option redirects the user to specify the settings into which the cleansed data must be moved.

viii) Click the 'Export Steps to data store' icon using the Navigation Pane.



• Steps when the 'Elastic' option is selected.



- i) Select the 'Elastic' option and provide the following details.
- ii) Data-Store Metadata Name: Provide a name for the data store metadata.
- iii) Select Mapping Id: Select a matching column from the drop-down menu.
- iv) Click the 'Export' option.

Ex	port	×
	Elastic	^
2	Data-Store Metadata Name Data Prep Data Store	
3	Select Mapping Id age	
		4 Export

v) A Success message appears to confirm.



vi) The settings get exported to the selected Elastic Settings. The user can see it under the Data Store Meta Data list of the Data Center module.

🗮 Data Center					New
Home	Search Data-Store Meta Data	Showing 486 out (ut of 486	
Data Connectors					
6	Data Prep Data Store	0	Ð	-	Î
Data Sets	pipeline logs viz2	0	Ð	1	Ĩ
🍙 Data Stores	XGBoost_Out	0	Ð	1	Î
😜 🛛 Data Store Meta Data	XGBoost_out	0	0	1	Î
🚉 Data Sheets	miphi	0	Ð	/	Î

- Steps when the 'Export as Scala Script' option is selected.
 - i) Select the 'Export as Scala Script' option and provide the following details.
 - ii) Enter the name of the Script.
 - iii) Click the '**Ok**' option.



Export	×
Elastic	~
1 Export As Scala Script.	<u>^</u>
2 Enter the name Data Prep Scala Script	
	3 Ok

iv) A success message appears, and the Scala Script gets downloaded to the system.



7.2. Export Steps to Pipeline

This option provides an option to specify the name in which the steps/transforms created as part of cleansing must be exposed to the pipeline module of the platform.

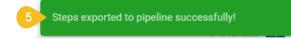
i) Click the 'Export Steps to Pipeline' icon.



- ii) The 'Export Steps to Pipeline' window opens.
- iii) Provide the name for the Data Prep script.
- iv) Click the 'Ok' option.
- v) A success message appears to assure the step.

2 Export Steps To Pipeline	×
3 Enter the name Data Prep Steps	
	Cancel Ok

vi) The Data Prep Script gets listed under the Data Scripts section of the Data Pipeline plugin.





7.3. Other options in the Navigation Pane

- **1.** Undo **?** : The user can Undo a list of the last few transforms. This button gets enabled only if we have applied at least one transform on the data.
- 2. Redo C: Redo a list of last few transforms, that was undone. If we have not undone any transform, then the 'redo' icon gets disabled.
- **3. Replay dirty** The '**Replay Dirty**' option when applied on the data from a specific step it replays all the transforms which are listed after the selected transform in the list of steps.
 - The '**Replay Dirty**' option gets enabled only when the user edits some transform step using the '**Steps**' tab.
 - To indicate what all transform steps will be affected, the listed steps get colored in red.
 - After the '**Replay Dirty**' function gets applied, all the steps that were colored in red become black and all the transforms get applied to the dataset.
- 4. **Replay All** E: The Replay All option allows the user to resample the data and replay the steps on the new data sample. It is useful when there is a change in the underlying dataset. It updates the data in the grid applying all the steps (In case of edit or steps added after edit).
 - Click the '**Replay All**' icon from the navigation pane.
 - The 'Replay All' window appears.
 - Select the 'Resampling & Replay' option using the checkbox (if required).
 - Click the '**Ok**' option.

le na	Replay All			×
le le le	Resampling & Replay			re
le	1 This might take a while.			
le	DUD	4	Cancel	

4. **Close the Preparation**: The user can exit from the preparation window and reach the landing page of data preparation.

Note: The standalone version of data preparation provides an option to export the prepared data to elastic so that that visualization modules can consume it.

8. Signing Out

The users can Sign-out from the Data Preparation tab at any given stage, but preferable is that the users should complete all the preparation tasks they wish to perform and save it before closing the tab or singing out from the Platform.

The Signing Out process for the Data Preparation has two steps:



1. Closing the BDB Data Preparation

Once you have completed the Data Preparation tasks, save your work and close the Data Preparation tab.

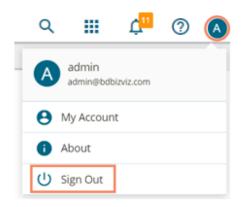
Click the **'Close'** button (the 'X' on the right edge) from the Data Preparation tab.



2. Sign Out from the BDB Platform

The following steps describe how to Sing-off from the BDB Platform.

- i) Click the 'User Profile' icon on the Platform homepage.
- ii) Click the 'Sign Out' option.



iii) The user successfully signs off from the BDB Platform.

Note:

- a. By clicking the 'Sign Out' option, the user gets back to the Sign-in page of the BDB platform.
- b. Click the 'About' option to open the default homepage for the BDB Platform.

BBB®			९ 🏭 🛱 💿 🖪
My Documents			Administration User
Welcome to BDB Decision platform BDB is a complete decision platform for all your business needs. Drive from data to dynamic visuals and derive an actionable insight into your business data.	Dashboard Designer Design, save and publish a splendid governed dashboards. Display relevant KPIs through comprehensive and stunning visual reports to attain your business objectives.	Business Story Business Story Go beyond the classic BI with our ground-breaking self-service BI tool. Gain pertinent insights into your business by creating wide-ranging views on your own without external IT help.	My Account About Sign Out learning plan your next business move. Access and apply accurate and customizable Predictive models to maximize future opportunities.
Avail 360° view of your business by assembling, processing, and analyzing the acquired data. Access incomparable analytics at any time from anywhere on any device. Version: 6.0.0	Data Center Data Center Supports a wide range of Data sources starting from the spreadsheets in your system to a cloud-based database. Establish connections to these data sources and build Data Sets or Data Stores to enable rich business intelligence	a wide range of Data a wide range of Data s starting from the est starting from the est in your system to a ed database. Establish s to these data sources s to these data sources s tandards to deliver the output in a bata Sets or Data Stores presentation-ready format.	



8.1. Forgot Password Option

The users are provided with a choice to change the password on the Login page of the platform.

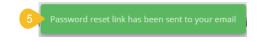
i) Click the 'Forgot Password?' option from the Sign In page.

Decision Platform
≌ Email / User Id *
• Password *
Auth Type Enterprise
Sign In
Convright © 2015-2021 BDR (BirViz Technologies Pur Ltd)

- ii) The 'Forgot Password?' page opens.
- iii) Provide the email id that is registered with BDB to send the reset password link.
- iv) Click the 'Continue' option.

Decision Platform
Porgot Password?
Please enter the registered email address to reset your password.
3 Email * admin.user@bdb.ai
Sign in
Continue
Copyright © 2015-2021 BDB (BizViz Technologies Pvt Ltd)

v) The user may be redirected to select a space in case of multiple spaces under one server link (The user needs to select a space and click the 'Continue' option once again). In case if a user does not have multiple spaces, a message appears to notify the user that the password reset link (The users receive the reset link via their registered email.)



vi) Click the link from your registered email.



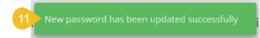
This email is sent in response to your request for a password reset.

6	Click on the below link to reset your Password Reset Password
	If you feel you have received this email in error or have any questions, please contact us at <u>support@bdbizviz.com</u>
	Thanks And Regards,
	Support Team

- vii) The user gets redirected to the 'Reset Password' page to set a new password.
- viii) Set a new password.
- ix) Confirm the newly set password.
- x) Click the '**Continue**' option.

	Reset Password
	You have confirmed ownership of the BDB account Please reset your password to get access.
8	New Password *
9	Confirm New Password *

xi) The password for the selected BDB account gets reset and a message appears to inform the user.



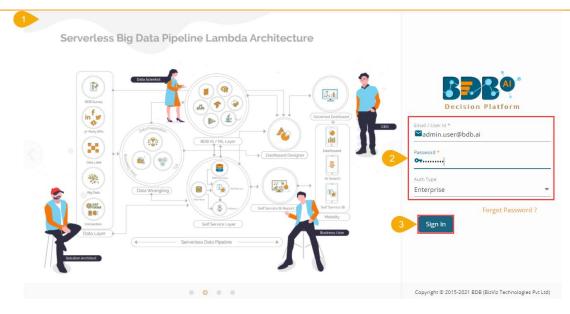
Note: The user gets redirected back to the Sign In page after successfully resetting the password.

8.2. Force Login

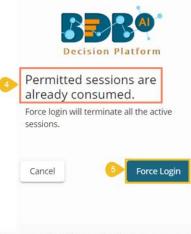
The '**Force Login**' functionality has been introduced to control the number of active sessions up to three. The users can access only 3 sessions at a time when they try to access the 4th session, a warning message displays to inform that the user has consumed the permitted sessions and, a click on the '**Force Login**' would kill all those active sessions.

- i) Navigate to the BDB Platform Login page.
- ii) Enter the valid credentials to log in.
- iii) Click the 'Sign In' option.





- iv) The user gets the following message if the permitted active sessions (3 sessions at a time) are consumed.
- v) Click the 'Force Login' option.



Copyright © 2015-2021 BDB (BizViz Technologies Pvt Ltd)

- vi) A warning message appears the currently active sessions get killed, and the user gets redirected to the BDB Platform Sign In page.
- vii) The user needs to provide valid credentials once again and click the '**Continue**' option to access the platform.