

Tableau

Module 12

Introduction to Tableau Prep



Agenda

Introduction to ETL work with Tableau Prep

- Introduction
- Connect data from different data bases for cleansing.
- Basics of creating workflows.
- How Tableau Prep helps quickly combine, join, shape, filter and clean data for analysis.

Preparation made easier

- Making data preparation simpler and accessible.
- Creation of smart examples with Tableau Prep.
- Create flows with Output in a desired format.

Integrating with Tableau Analytical workflow

- Getting deeper insights into the data with great visual experience.
- Understanding the seamless process from data preparation to analysis with Tableau Prep.

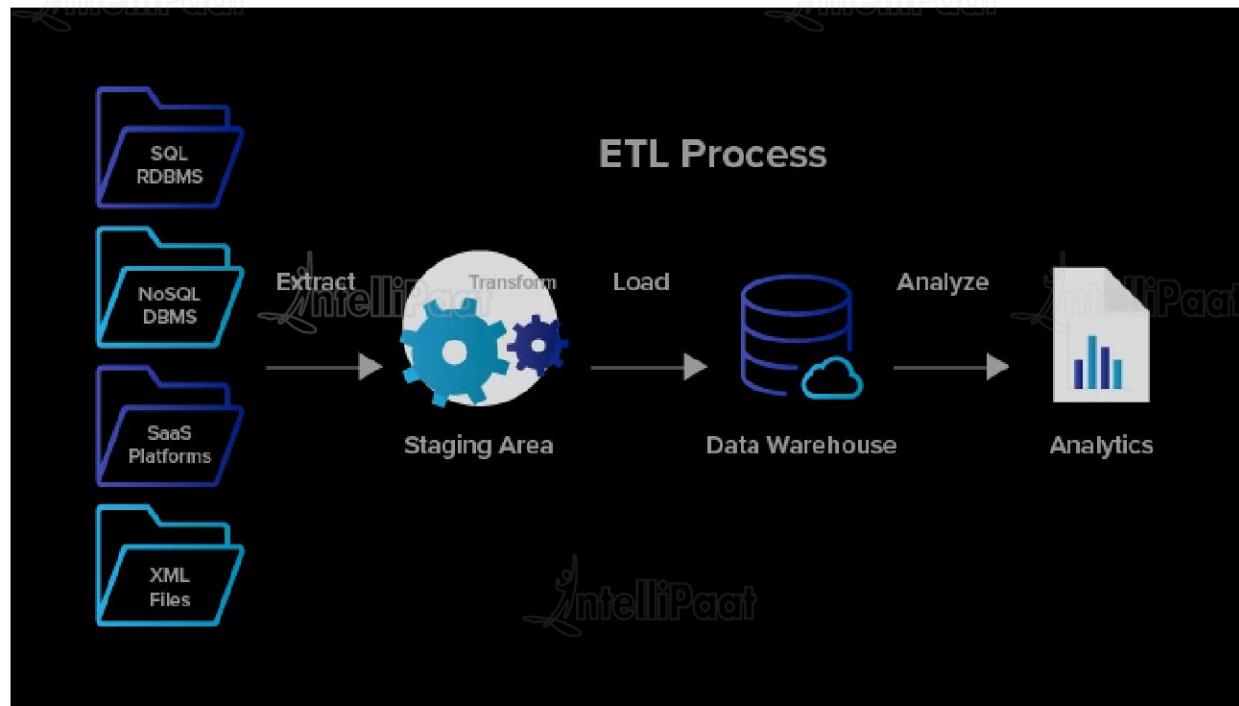


Introduction to ETL work with Tableau Prep



What is ETL?

- ETL stands for Extract, Transform, Load. It is a process that extracts the data from different source systems, then transforms the data (like applying calculations, concatenations, filters etc.) and finally outputs the data in a format with which you can proceed with the Analysis work in BI tools like Tableau Desktop.
- Tableau Prep is known as the ETL tool/preparation tool of Tableau.



What is Data Visualization?

- Data visualization is the process of converting raw data into easily understandable formats of information that enable fast and effective decision-making. Charts, diagrams, pictures, etc. are examples of data visualization.

A	B	C	D	E	F	G	H	I	J	K	
1	Iga_id	pop_2006	Iga	state	zone	area_sq_km	surveying_ef	pop_density	unique_iga	longitude	latitude
2	1	100180	Abadam	Borno	Northeast	3967.05423	Others	25.2529948	borno_abada	13.3256741	13.4282932
3	2	58444	Abaji	FCT	North-Centr	992.36124	113	58.8938761	fct_abaji	6.85341375	8.87756561
4	3	139090	Abak	Akwa Ibom	South-South	189.984563	Others	732.12113	akwa_ibom	7.7722792	5.01988684
5	4	151723	Abakaliki	Ebonyi	Southeast	584.055167	Others	259.775118	ebonyi_abak	8.21988187	6.2586811
6	5	107488	Aba North	Abia	Southeast	22.7750611	Others	4719.54826	abia_aba_no	7.37164362	5.10690037
7	6	423852	Aba South	Abia	Southeast	49.1665418	Others	8620.74053	abia_aba_so	7.35065009	5.08134569
8	7	201329	Abeokuta Nc Ogun	Ogun	Southwest	808.831493	Others	248.913404	ogun_abeok	3.1735877	7.22300486
9	8	250278	Abeokuta So Ogun	Ogun	Southwest	71.4715348	Others	3501.78572	ogun_abeok	3.35694358	7.15311807
10	9	144802	Abi	Cross River	South-South	281.772825	Others	513.896257	cross_river_i	8.0266891	5.89308847
11	10	195652	Aboh Mbaise Imo	Imo	Southeast	184.48387	Others	1060.53716	imo_aboh_ir	7.25224758	5.43261623
12	11	282988	Abua Odual Rivers	Rivers	South-South	704.338115	Others	401.77862	rivers_abua	6.56209631	4.83853956
13	12	202194	Adavi	Kogi	North-Centr	718.503598	Others	281.409864	kogi_adavi	6.44459043	7.67423257
14	13	308621	Ado-Ekiti	Ekiti	Southwest	293.6513	Others	1050.97781	ekiti_ado_e	5.27997457	7.61042627
15	14	526565	Ado-Odo/Ot: Ogun	Ogun	Southwest	878.916456	148	599.106998	ogun_ado_o	3.08462514	6.1884102
16	15	134173	Afijio	Oyo	Southwest	722.693666	Others	185.656809	oyo_afijio	3.90125509	7.74211826
17	16	156611	Afikpo North Ebonyi	Ebonyi	Southeast	240.157467	Others	652.117972	ebonyi_afikp	7.9233965	5.87842851
18	17	157072	Afikpo South Ebonyi	Ebonyi	Southeast	377.675196	113	415.891755	ebonyi_afikp	7.81363734	5.83719147
19	18	132907	Agadez	Niger	North-Centr	1904.35471	Others	69.7910949	niger_agadez	6.44045921	8.94623535
20	19	115523	Agatu	Benue	North-Centr	1014.60843	Others	113.859689	benue_agatu	7.88215837	7.86983845
21	20	57413	Agwara	Niger	North-Centr	1539.669	148	37.2891837	niger_agwara	4.52764087	10.7410218
22	21	459939	Agege	Lagos	Southwest	11.214557	Others	41012.6765	lagos_agege	3.31600267	6.62357986
23	22	370172	Aguata	Anambra	Southeast	195.102574	Others	1897.31992	anambra_aguata	7.0880433	5.98862094
24	23	170902	Ahiazu-Mba Imo	Imo	Southeast	113.716283	Others	1502.88063	imo_ahiazu_u	7.26418199	5.56020001
25	24	166747	Ahoadha East Rivers	Rivers	South-South	341.173834	Others	488.744984	rivers_ahoadha	6.6410198	5.06535919
26	25	249425	Ahodata West Rivers	Rivers	South-South	403.45939	Others	618.215875	rivers_ahodata	6.51660569	5.05176308
27	26	122321	Ajaokuta	Kogi	North-Centr	1362.39948	113	89.7835045	kogi_ajaokut	6.5963902	7.51474121
28	27	684105	Ajegromi-Ifelegba	Lagos	Southwest	12.3493909	148	55395.8493	lagos_ajegromi	3.3374815	6.45356193
29	28	174137	Ajingi	Kano	Northwest	713.799255	Others	243.957946	kano_ajingi	9.06909277	11.9731313
30	29	151125	Akamkpa	Cross River	South-South	5001.71991	Others	30.2146067	cross_river_i	8.53398001	5.40945643
31	30	211359	Akinyele	Oyo	Southwest	518.870311	Others	407.344563	oyo_akinyele	3.93139439	7.56243229

Raw Data

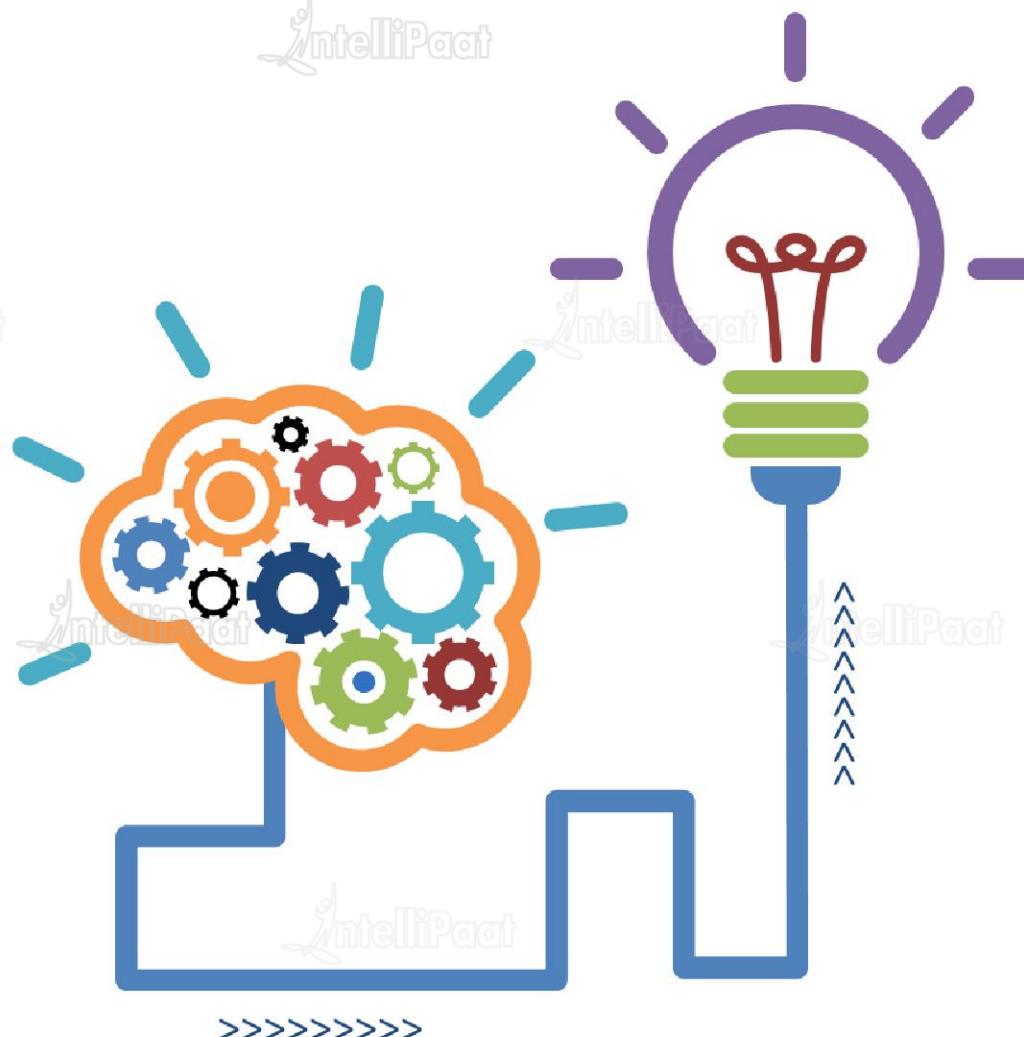


Converted to Meaningful Visual Information

A Picture is Worth a Thousand Words



**Humans are very
good in solving
complex problems.**



**Visual representations
enhance the human
cognitive process.**

Business Needs



Businesses realize that a visual can communicate:

- More information than a table
- In a much smaller place

DATA VISUALIZATION



Data visualization leads to better:

- Understanding
- Business decisions, accurately and quickly

WE UNDERSTAND
YOUR NEEDS

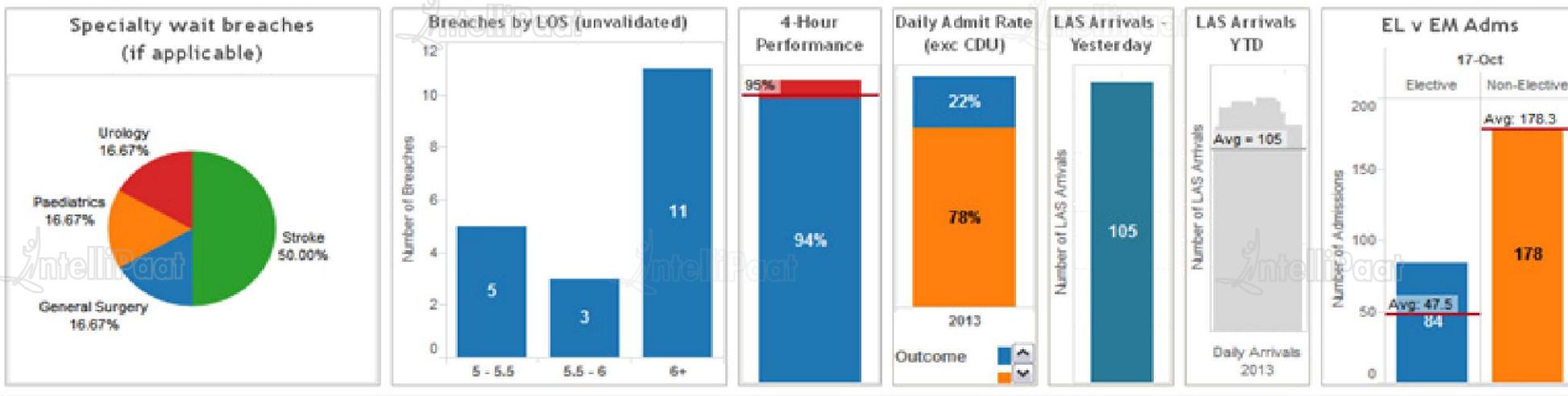
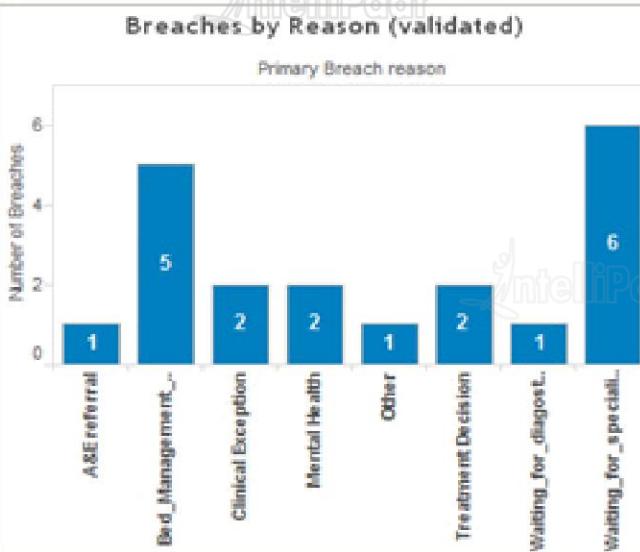
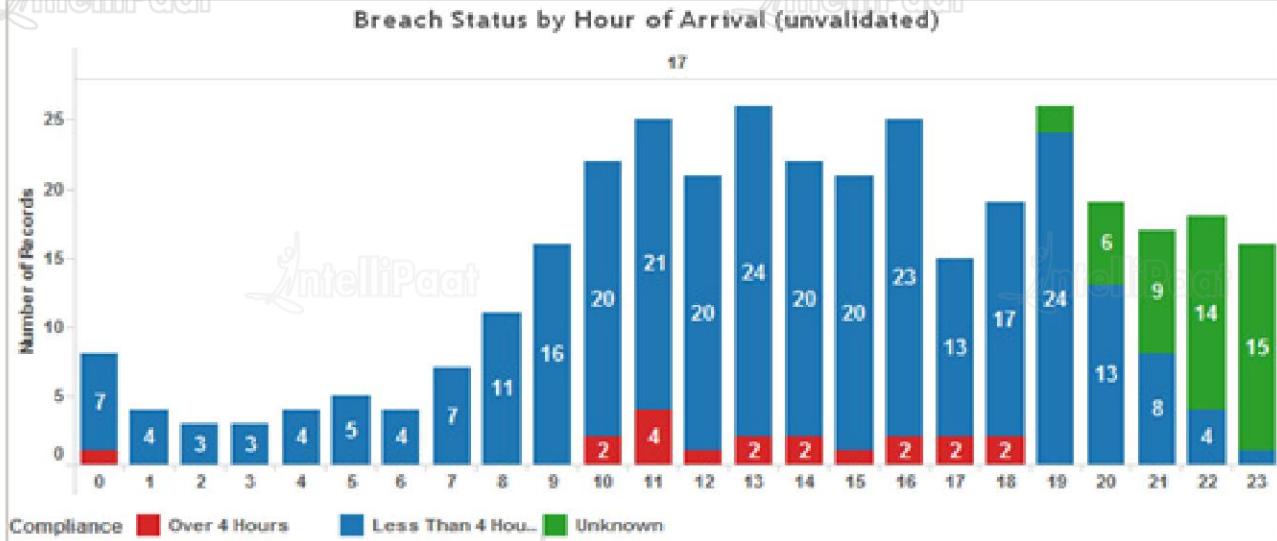
Appealing and dynamic visuals that:

- Manipulate and interact directly and easily with the data
- Can update automatically
- Facilitate data analytics and data discovery
- Identify trends and patterns
- Act on emerging trends fast
- Help in business decisions

Healthcare Dashboard



ED Dashboard - Daily



Introduction to Tableau Prep

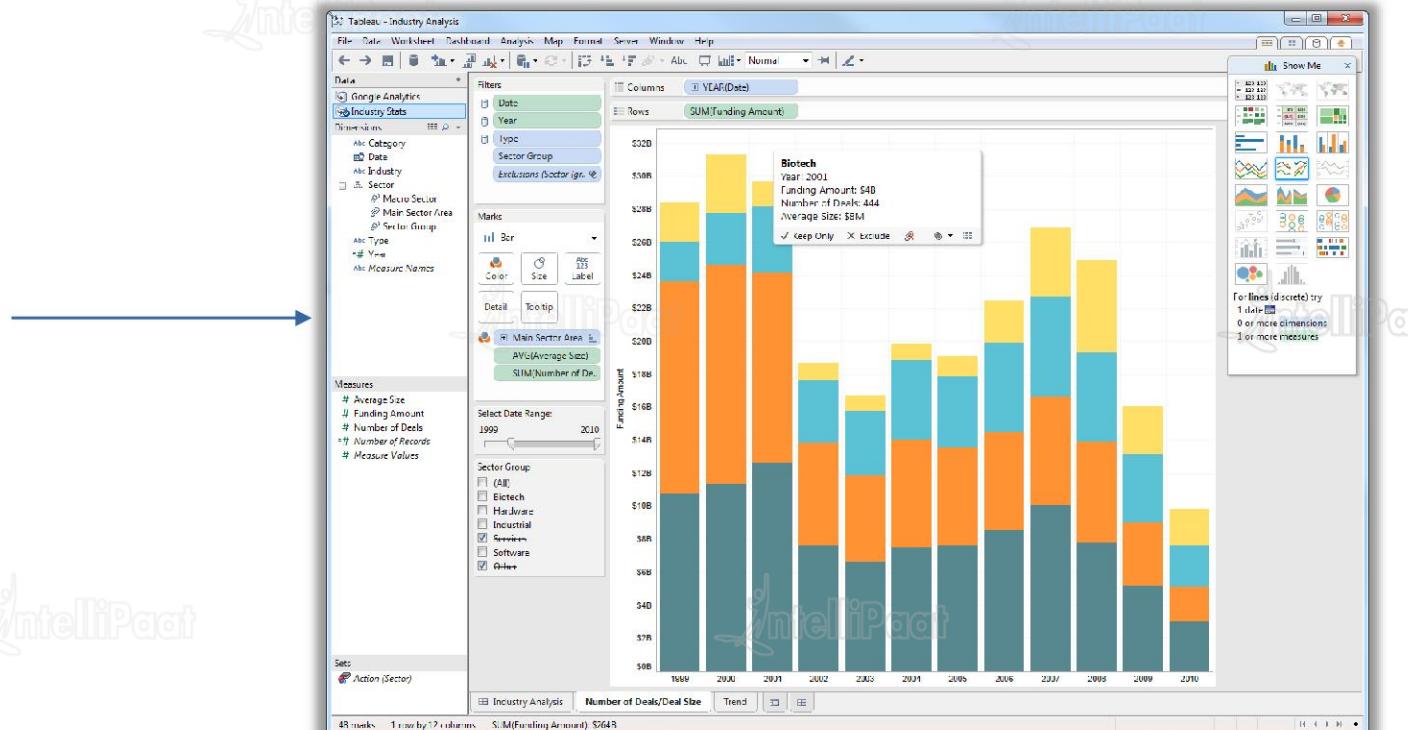
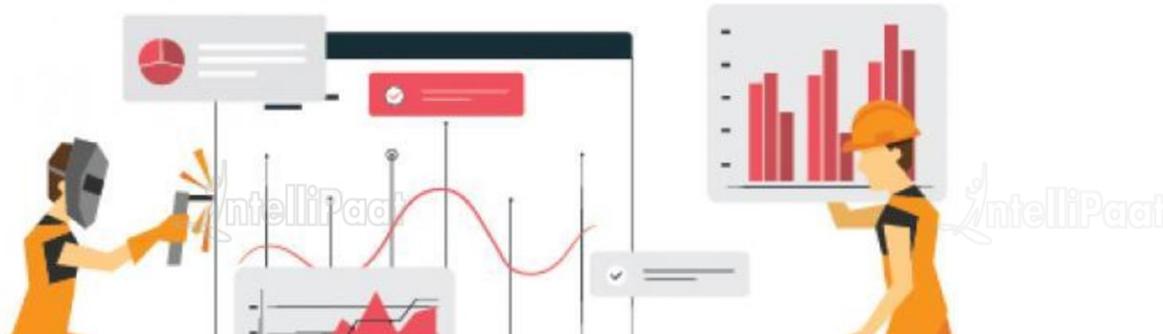


What is Tableau?

A Tableau is an interactive Business Intelligence (BI) software used for data preparation, data visualisation and data analytics. An American computer software company founded in January, 2003, by Chris Stolte developed this software.

TABLEAU DATA PREP

Cleaning and Restructuring
Data for Visualizations



Power of Tableau



Tableau Vs. Other Vendors



- Tableau is an industry leader in Business Intelligence and Analytics.
- Tableau is recognized as a leader in Gartner Magic Quadrant for Business Intelligence and Analytics Platforms for the 8th consecutive year.

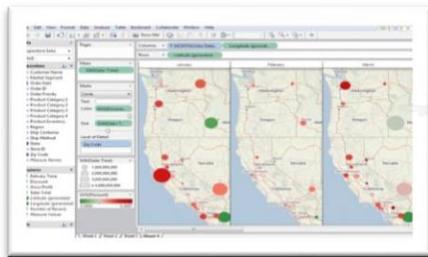


Tableau Products



Tableau Desktop

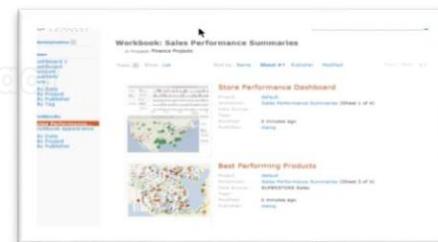
Create



- Ad hoc analytics, dashboards, reports and graphs
- Explore, visualize and analyze data
- Create dashboards to consolidate multiple views
- Deliver interactive data experiences

Tableau Server

Share - Web



- Used for on-premises deployments.
- Business Intelligence solution scales to organizations of all sizes
- Share visual analytics with anyone with a web browser
- Publish interactive analytics or dashboards
- Secure information and manage metadata
- Collaborate with others

Tableau Reader

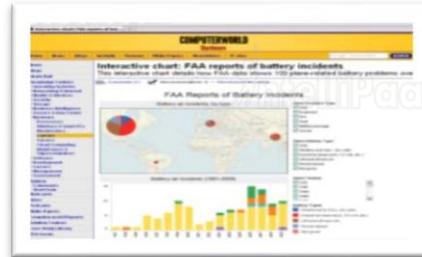
Share - Local



- Share visualizations and dashboards on the desktop
- Filter, sort and page through the views
- “Acrobat for Data”
- Free download

Tableau Public

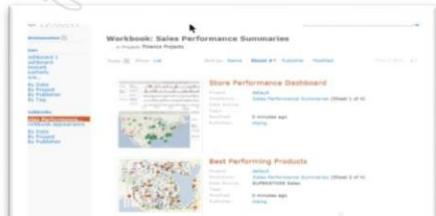
Share - Everyone



- Create and publish interactive visualizations and dashboards
- Embed in websites and blogs
- Free download and free hosting service

Tableau Online

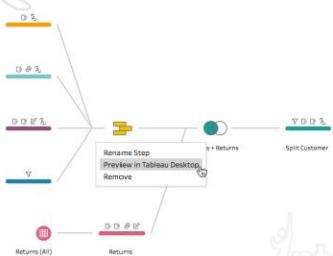
Share - Cloud



- Used for cloud based deployments.
- Managed and maintained by Tableau Company.
- High level of security.
- Publish interactive analytics or dashboards
- Secure information and manage metadata
- Collaborate with others

Tableau Prep

ETL Tool of Tableau



- Extract-Transform-Load Tool.
- Used for preparation and modification work on data before creating dashboards with Tableau desktop or Public.

How Does Tableau Prep Work?

- Tableau is based on three simple concepts:

Connect

Connect Tableau to any data source that you want to prepare and analyze.

Prepare

Prepare/Cleanse the data the way you want, such as filter, sort, perform aggregations on data, summarize and so on.

Analyze and Share

You can analyze and share results with others by:

- Create Dashboards using Tableau Desktop.
- Sharing workbooks with other Tableau users.
- Printing to PDF
- Using Tableau Server to publish or embed your views across organization.



How to Download and Install Tableau prep

Connecting to Data



Click the Tableau icon on the upper left-hand corner of any page.



Under Connect, you can:

- Connect to data stored in a file, such as Microsoft Excel or Access.
- Connect to data that is stored on a server, such as Tableau Server, Microsoft SQL Server or Google Analytics.
- Connect to a data source that you've connected to before.
- Tableau supports the ability to connect to wide ranging data stored in a wide variety of places.

You can also open an existing flow that you had already created – by clicking on Open workflow.

A screenshot of the Tableau Prep Builder software interface. The window title is 'Tableau Prep Builder - Flow1'. On the left, there's a sidebar titled 'Connections' with a list of options: 'To a File' (PDF file, Microsoft Excel, Statistical file, Tableau extract, Text file), 'To a Server' (Tableau Server, Amazon Athena, Amazon Aurora, Amazon EMR Hadoop Hive, Amazon Redshift, Apache Drill, Aster Database, Azure SQL Data Warehouse, Cloudera Hadoop, Denodo, EXASOL), and two specific flows named 'Superstore' and 'Worldindicators'. The main workspace shows a small portion of a data flow with some nodes and connections.

Tableau Prep - Workflow



IntelliPaat

Orders (East)

Orders (West)

Orders (Central)

Orders (South)

All Orders

Orders + Returns

Split Customer

Create 'All Orders'

Quota

Pivot Quotas

Aggregate

Quota + Orders

Create 'Annual ...'

22 Fields 2K Rows

Changes (2)

State 11

Row ID 2K

Order ID 822

Segment 3

Customer ID 512

Customer Name 512

Ship Mode 4

Order Date 604

01/01/2015, 1... 01/01/2019, 1...

Sales	Quantity	Profit	Discount	Region	State	Row ID	Order ID	Segment	Customer ID	Customer Name	Ship Mode	Order Date	Ship Date
18.648	7	-12.432	0.7	South	North Carolina	231	US-2015-156216	Corporate	EA-14035	Erin Ashbrook	Standard Class	09/13/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
178.384	2	22.298	0.2	South	Florida	315	CA-2015-167850	Corporate	AG-10525	Andy Gerbode	Standard Class	08/09/2015, 12:00:00 AM	08/16/2015, 12:00:00 AM
15.552	3	5.4432	0.2	South	Florida	316	CA-2015-167850	Corporate	AG-10525	Andy Gerbode	Standard Class	08/09/2015, 12:00:00 AM	08/16/2015, 12:00:00 AM
39.072	6	9.768	0.2	South	North Carolina	404	CA-2015-155208	Corporate	SP-20650	Stephanie Phelps	Standard Class	04/16/2015, 12:00:00 AM	04/20/2015, 12:00:00 AM
10.368	2	3.6288	0.2	South	North Carolina	705	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
166.84	5	18.7695	0.2	South	North Carolina	706	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
15.216	1	2.2824	0.2	South	North Carolina	707	CA-2015-138527	Corporate	BN-11470	Brad Norvell	Standard Class	09/12/2015, 12:00:00 AM	09/17/2015, 12:00:00 AM
11.36	2	5.3392	0	South	Louisiana	764	CA-2015-162775	Corporate	CS-12250	Chris Selesnick	Second Class	01/13/2015, 12:00:00 AM	01/15/2015, 12:00:00 AM

Using Data Interpreter



Connections

Add

little_crime1_excel
Excel

little_crime2_excel
Excel

Sheets

Use Data Interpreter
Data Interpreter might be able to clean your Excel workbook.

little_crime2_excel

Click on
Review the results

Cleaned with Data Interpreter

[Review the results.](#) (To undo changes, clear the check box.)

little_crime2_excel

little_crime2...cel A1:G10000

An Excel file is **generated** with the interpreted null values.
Look at the 'A's!

A	B	C	D	E	F	G
ID	Psa	Dc_Key	Location_IUCR_Gen	Text_Gen	Police_Dis	
2	0	1	2,02E+11 800 BLOCK	600 Thefts	13	
3	1	3	2,02E+11 1500 BLOC	600 Thefts	16	
4	2	3	2,02E+11 1500 BLOC	600 Thefts	16	
5	3 A		2,02E+11 GATE C-18	600 Thefts		
6	61	1	2,02E+11 900 BLOCK	600 Thefts	3	
7	62	1	2,02E+11 100 BLOCK	600 Thefts	20	
8	63	2	2,02E+11 5200 BLOC	600 Thefts	20	
9	64	1	2,02E+11 N 17TH ST	800 Other Assaults		
10	65 A		2,02E+11 PIA GATE	800 Other Assaults		
11	66	1	2,02E+11 100 BLOCK	600 Thefts	20	
12	67	4	2,02E+11 1400 BLOC	600 Thefts	16	
13	68	2	2,02E+11 5900 BLOC	2600 All Other	15	
14	69	2	2,02E+11 1100 BLOC	600 Theft from	11	
15	70 A		2,02E+11 PIA GATE	600 Thefts		
16	71 A		2,02E+11 GATE E-8	600 Thefts		
17	72 A		2,02E+11 PIA TERMI	600 Thefts		
18	73 A		2,02E+11 98 PIA WA	600 Thefts	22	
19	74 A		2,02E+11 GATE A-7	600 Thefts		
20	75	2	2,02E+11 400 BLOCK	200 Rape	4	
21	76	2	2,02E+11 1200 BLOC	200 Rape	5	
22	77	2	2,02E+11 3200 BLOC	1100 Fraud	7	
23	78	2	2,02E+11 2800 BLOC	600 Thefts	16	
24	79	2	2,02E+11 5200 BLOC	800 Other Ass	20	
25	80	1	2,02E+11 N 410 N 6	300 Robbery No Firearm		

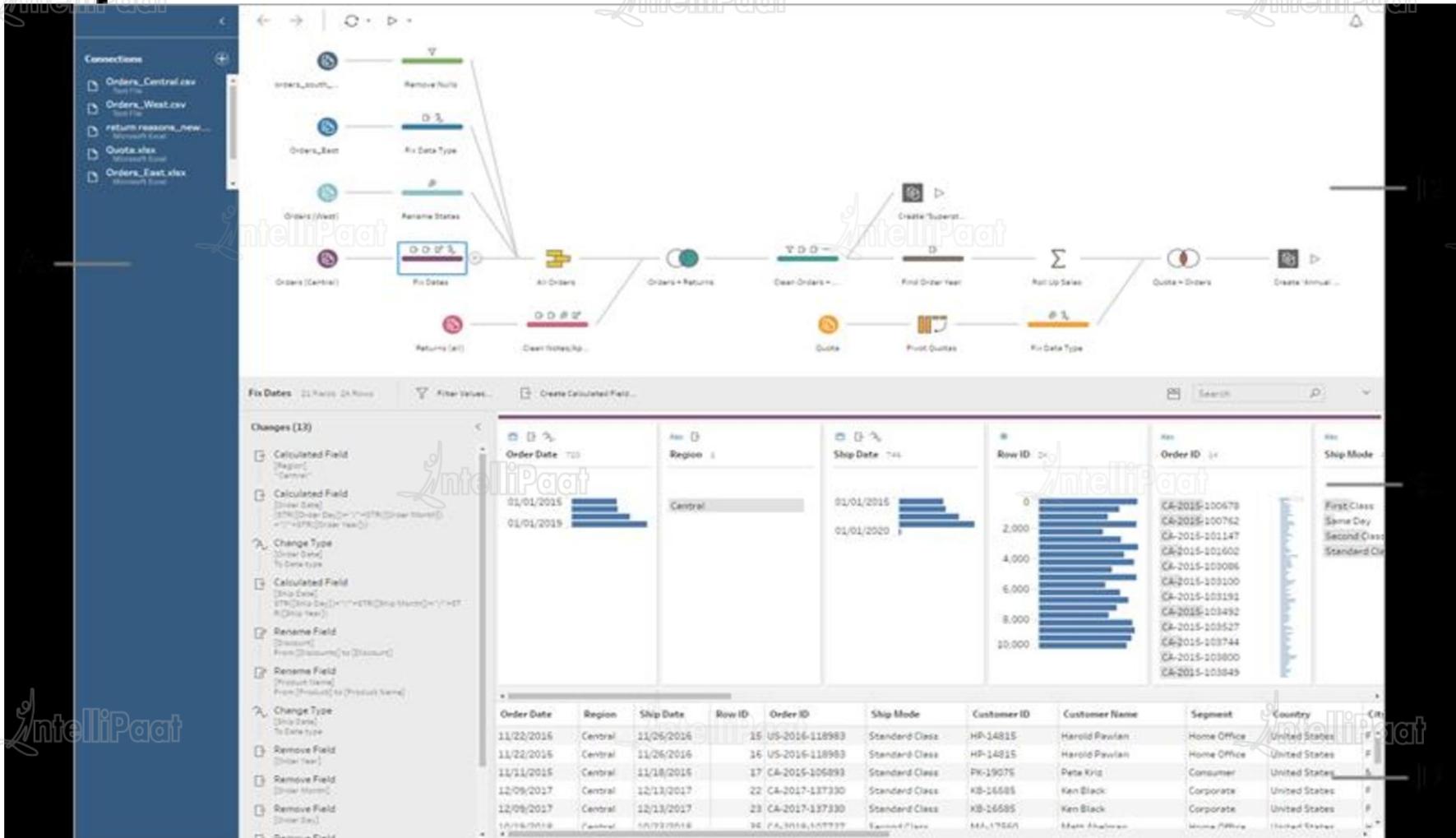


Preparation made Easier



How to start with Tableau Prep?

A tour of the Tableau Prep workspace



A tour of the Tableau Prep workspace

The Tableau Prep workspace consists of the **Connections pane (A)** where you connect to your data sources, and three coordinated areas that help you interact with and explore your data:

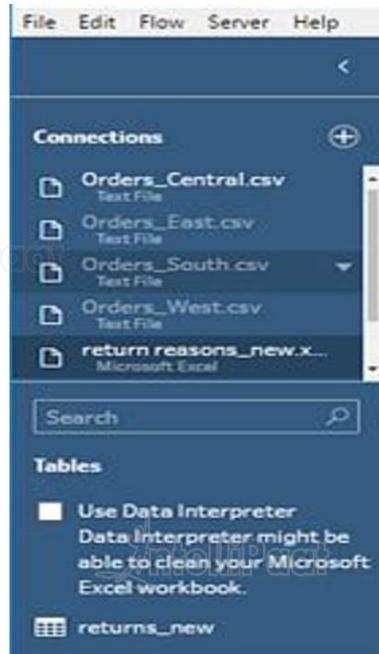
Flow pane (B): A visual representation of your operation steps as you prepare your data.
This is where you add steps to build your flow.

Profile pane (C): A summary of each field in your data sample. See the shape of your data and quickly find outliers and null values.

Data grid (D): The row level detail for your data.

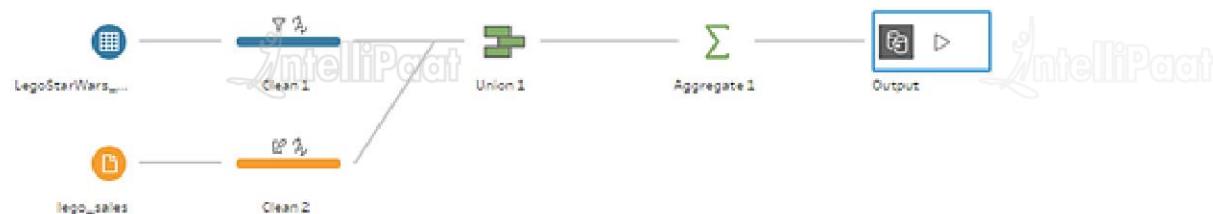
Connections pane

- On the left side of the workspace is the Connections pane, which shows the databases and files you are connected to. Add connections to one or more data sources and then drag the tables you want to work with into the Flow pane.



Flow pane

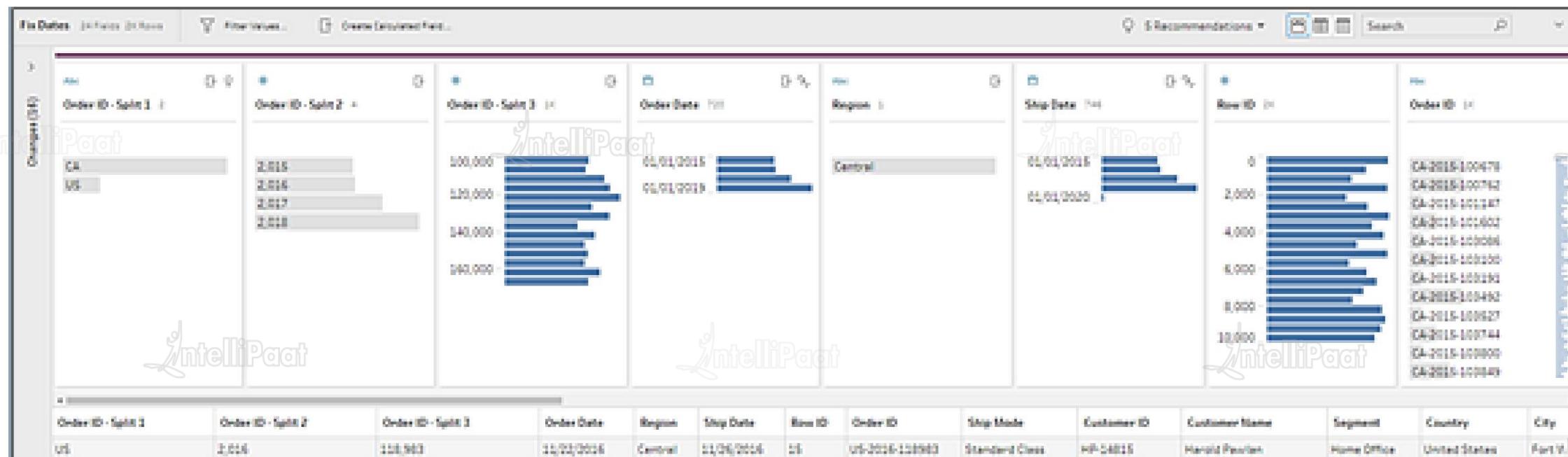
- At the top of the workspace is the Flow pane.
- This is where you will build your flow. As you connect to, clean, shape, and combine your data, steps appear in the Flow pane and align from left to right along the top.
- These steps tell you what kind of operation is being applied, in what order, and how your data is affected by it. For example, the Join step shows you which join type you've applied, the join clauses, recommended join clauses, and the fields of the tables that are included in the join.



- You start your flow by dragging tables into the Flow pane. Here you can add additional data sets, pivot your data, union or join data, create aggregations, and generate output files in the form of .tds files, Hyper extract (.hyper) files, or published data sources that you can use in Tableau. You can even write your output data to a database.

Profile pane

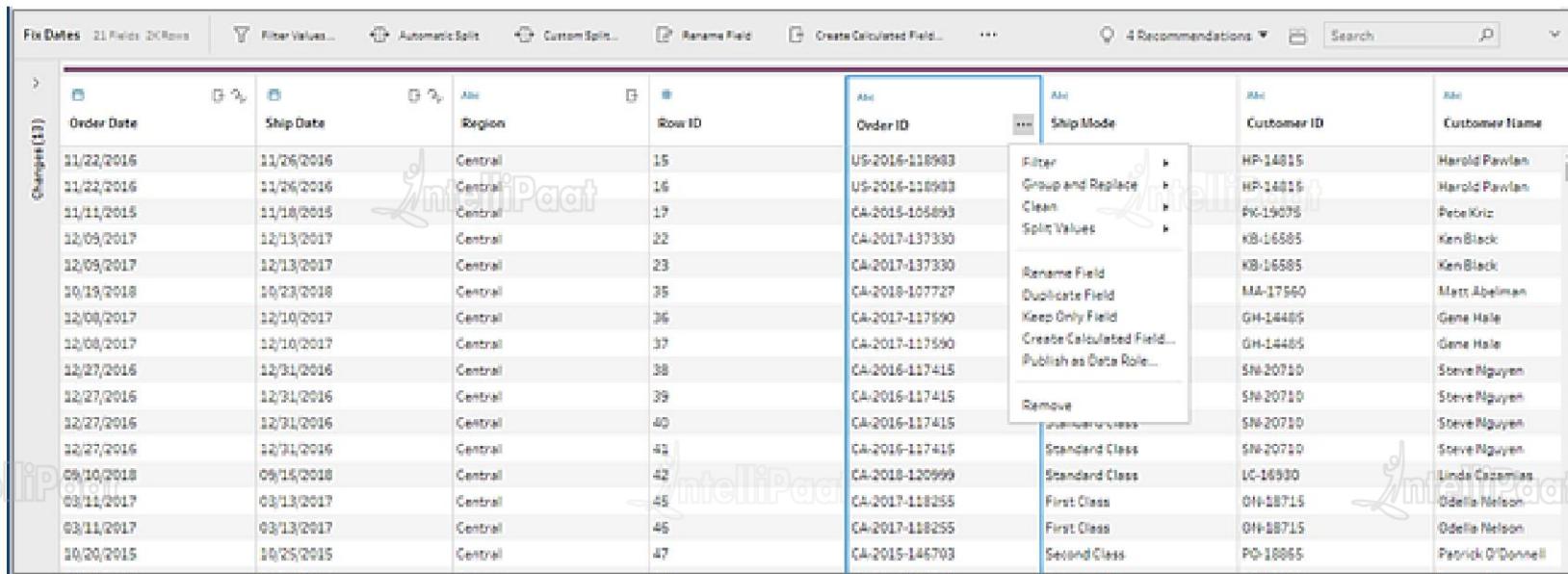
- In the center of the workspace is the Profile pane.
- The Profile pane shows you the structure of your data at any point in the flow.
- The structure of your data can be represented in different ways depending on the operation you want to perform on your data or the step that you select in the Flow pane.



- At the top of the Profile pane is a toolbar that shows you the cleaning operations that you can perform for each step in your flow. An options menu also appears on each card in the Profile pane where you can select the different operations that you can perform on the data.
- For example:
- Search, sort, and split fields
- Filter, include, or exclude values
- Find and fix null values
- Rename fields
- Clean up data entry errors using group values or quick cleaning operations
- Use automatic data parse to change data types
- Rearrange the order of your field columns by dragging and dropping them where you want them

Data grid

- At the bottom of the workspace is the Data grid, which shows you the row level detail in your data.
- The values displayed in the Data grid reflect the operations defined in the Profile pane.
- You can perform the same cleaning operations here as you can in the Profile pane if you prefer to work at a more detailed level.
- Click the Collapse Profiles icon on the toolbar to collapse (and expand) the Profile pane to see your options.



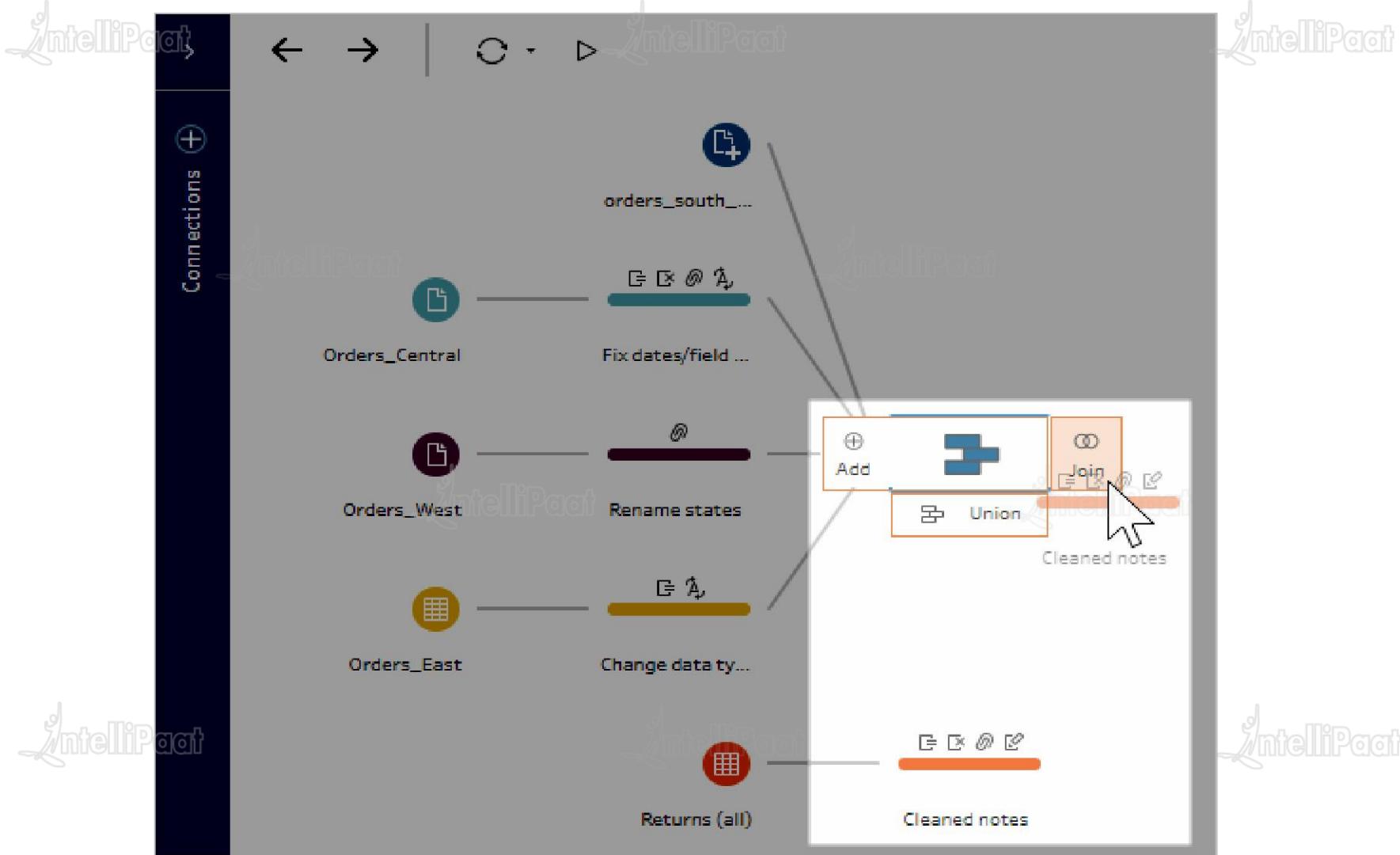
A screenshot of the Alteryx Data grid interface. The grid displays a table with columns: Order Date, Ship Date, Region, Row ID, Order ID, Ship Mode, Customer ID, and Customer Name. A context menu is open over the cell containing 'CA-2016-118983' in the Order ID column. The menu options include: Filter, Group and Replace, Clean, Split Values, Rename Field, Duplicate Field, Keep Only Field, Create Calculated Field..., Publish as Data Role..., and Remove. The 'Clean' option is highlighted with a blue border. The top of the screen shows the Alteryx ribbon with various tabs like Fix Dates, 21 Fields, 2K Rows, Filter Values..., Automatic Split, Custom Split..., Rename Field, Create Calculated Field..., 4 Recommendations, Search, and a Help icon.

Order Date	Ship Date	Region	Row ID	Order ID	Ship Mode	Customer ID	Customer Name
11/22/2016	11/26/2016	Central	15	US-2016-118983	Filter	HP-14815	Harold Pawlen
11/22/2016	11/26/2016	Central	16	US-2016-118983	Group and Replace	HP-14815	Harold Pawlen
11/11/2015	11/18/2015	Central	17	CA-2015-105893	Clean	WA-196975	Pete Kinc
12/09/2017	12/13/2017	Central	22	CA-2017-137330	Split Values	KB-165885	Ken Black
12/09/2017	12/13/2017	Central	23	CA-2017-137330		KB-165885	Ken Black
10/19/2018	10/23/2018	Central	35	CA-2018-107727		MA-17560	Marty Abelman
12/08/2017	12/10/2017	Central	36	CA-2017-117590		GH-14485	Gene Hale
12/08/2017	12/10/2017	Central	37	CA-2017-117590		GH-14485	Gene Hale
12/27/2016	12/31/2016	Central	38	CA-2016-117415		SN-20710	Steve Nguyen
12/27/2016	12/31/2016	Central	39	CA-2016-117415		SN-20710	Steve Nguyen
12/27/2016	12/31/2016	Central	40	CA-2016-117415		SN-20710	Steve Nguyen
12/27/2016	12/31/2016	Central	41	CA-2016-117415		SN-20710	Steve Nguyen
09/10/2018	09/16/2018	Central	42	CA-2018-120599		IC-16930	Linda Cizamias
03/11/2017	03/13/2017	Central	43	CA-2017-118255		OH-18715	Odella Nelson
03/11/2017	03/13/2017	Central	44	CA-2017-118255		OH-18715	Odella Nelson
03/20/2015	03/25/2015	Central	45	CA-2015-145703		PO-10065	Patrick O'Donnell

How Tableau Prep stores your data

- When you connect Tableau Prep to your data and create a flow, it stores the frequently used data in a .hyper file.
- For large data sets, this might be a sample of the data.
- Any stored data is saved under a secure temporary file directory in a file named Prep BuilderXXXXX, where XXXXX represents a universally unique identifier (UUID).
- After you save the flow, the file is deleted.

Workflow Examples

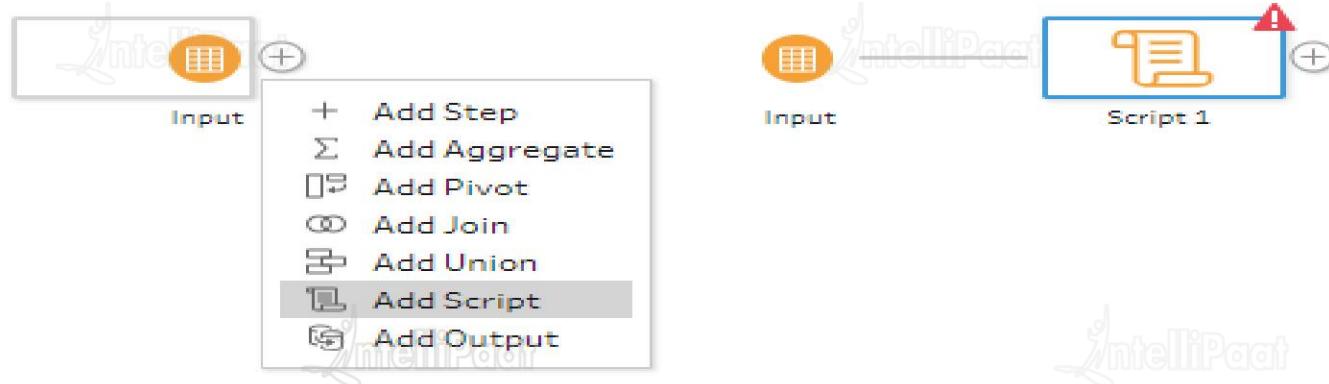




Integrating with Tableau Analytical workflow

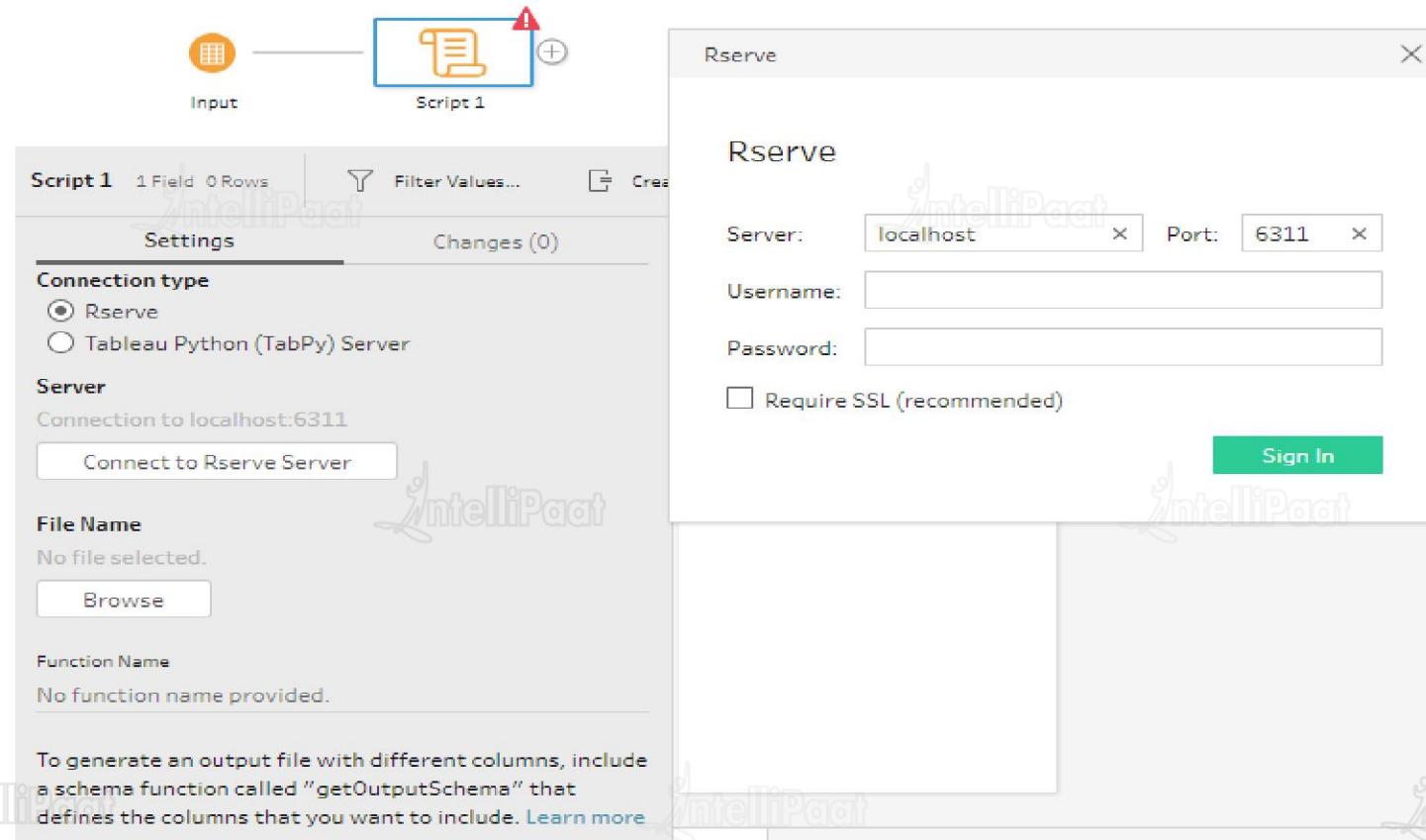
Getting Started with R Scripts in Tableau

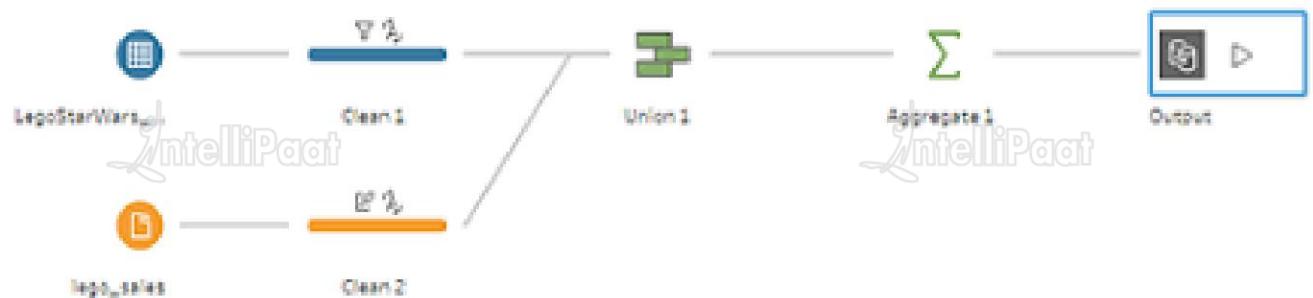
Prep



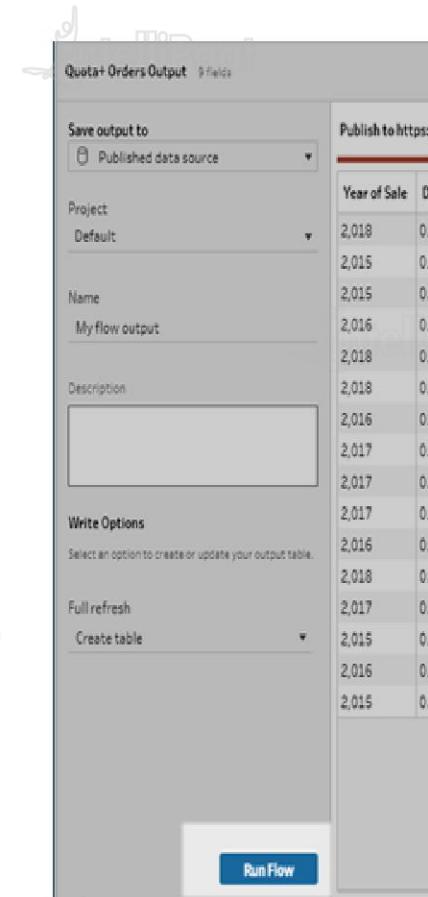
- This allows you to use an R or Python Script within your flow.

Connecting Rserve or TabPy with Prep





- Create and Run the output flow.



The screenshot shows the configuration of an output flow named "Quota+Orders Output".

Save output to: Published data source

Project: Default

Name: My flow output

Description: (empty)

Write Options: Select an option to create or update your output table.

Full refresh: Createable

Year of Sale	Order
2,018	0.1
2,015	0.1
2,015	0.1
2,016	0.1
2,018	0.1
2,018	0.1
2,016	0.1
2,017	0.1
2,017	0.1
2,017	0.1
2,017	0.1
2,016	0.1
2,018	0.2
2,017	0.2
2,015	0.1
2,016	0.2
2,015	0.2

Run Flow

- The transformed output file saved can now be connected to Tableau desktop for building dashboards and stories.



Quiz



Q 1. In which year Tableau Prep was introduced?

- a) 2020
- b) 2010
- c) 2018
- d) 2003



Quiz



Q 2. How is Tableau Prep defined?

- a) ETL TOOL
- b) Sharing Tool
- c) Developing Tool
- d) Analyzing Tool



Quiz



Q 3. Which is the latest product introduced by Tableau?

- a) Tableau Reader
- b) Tableau Prep
- c) Tableau Desktop
- d) Tableau Public



Quiz



Q 4. Tableau Public can also be used as an ETL Tool.

a) True

b) False



Quiz



Q 5. When was “Add Script” introduced in Tableau Prep?

- a) 2019.3
- b) 2018.1
- c) 2020.3.1
- d) 2017.2





Thank You