Introduction to Tableau Prep Builder

MRCB | s5 | 2019

Josep Curto | Professor, IE Business School | CEO, Delfos Research

SCHOOL OF HUMAN SCIENCES & TECHNOLOGY

Agenda

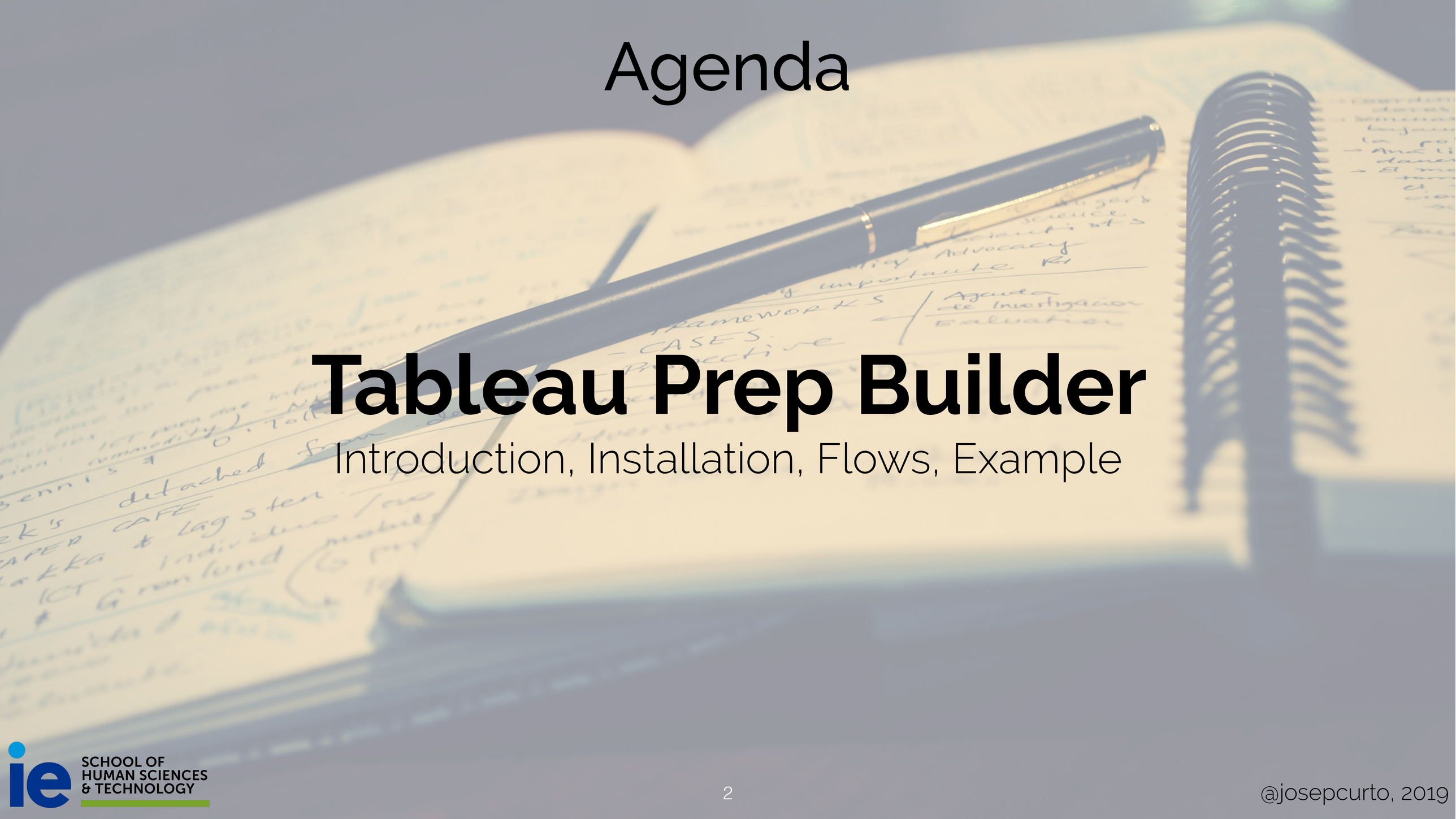


Tableau Prep Builder

Introduction, Installation, Flows, Example

Introduction

What is Tableau Prep Builder

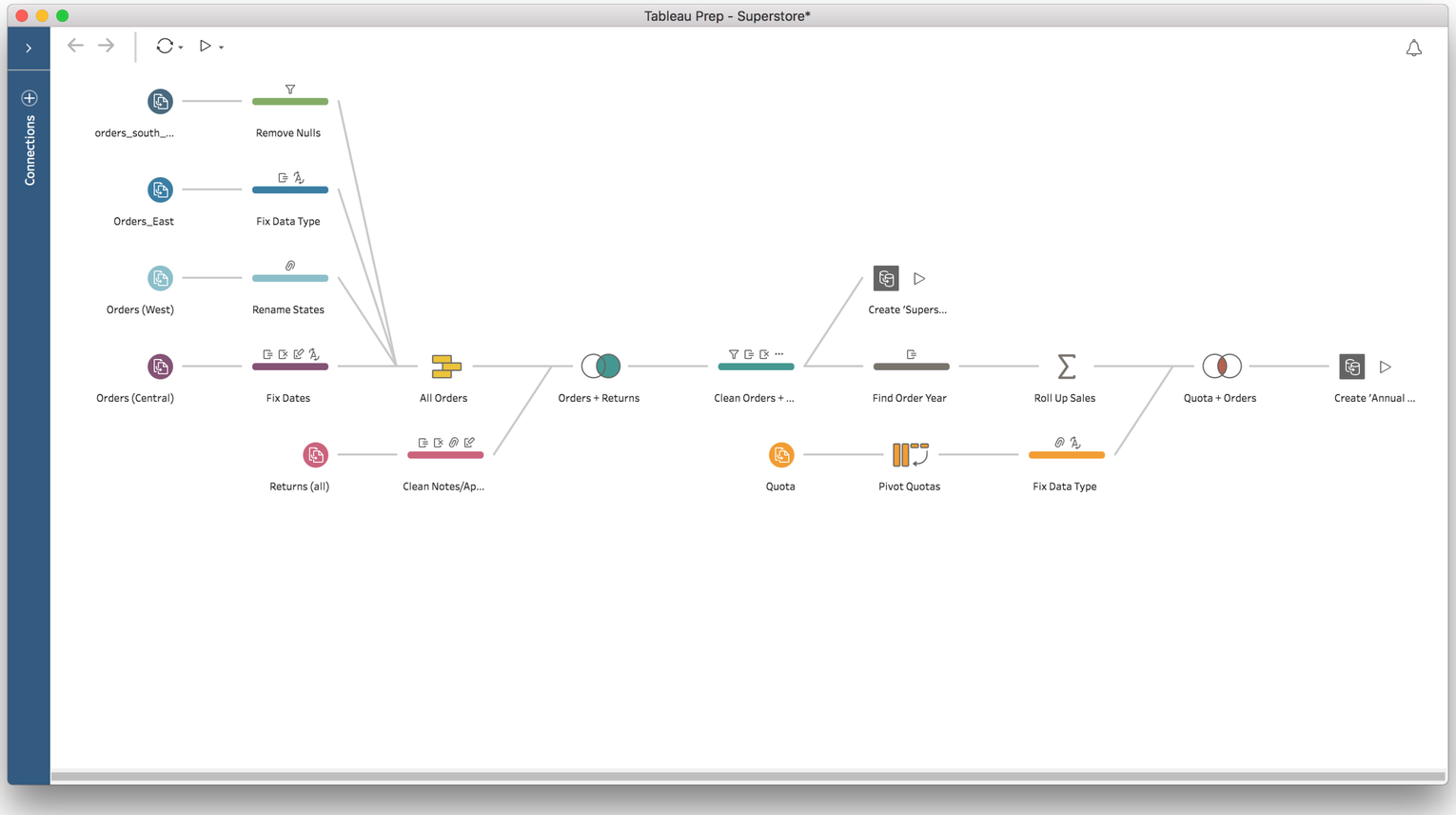
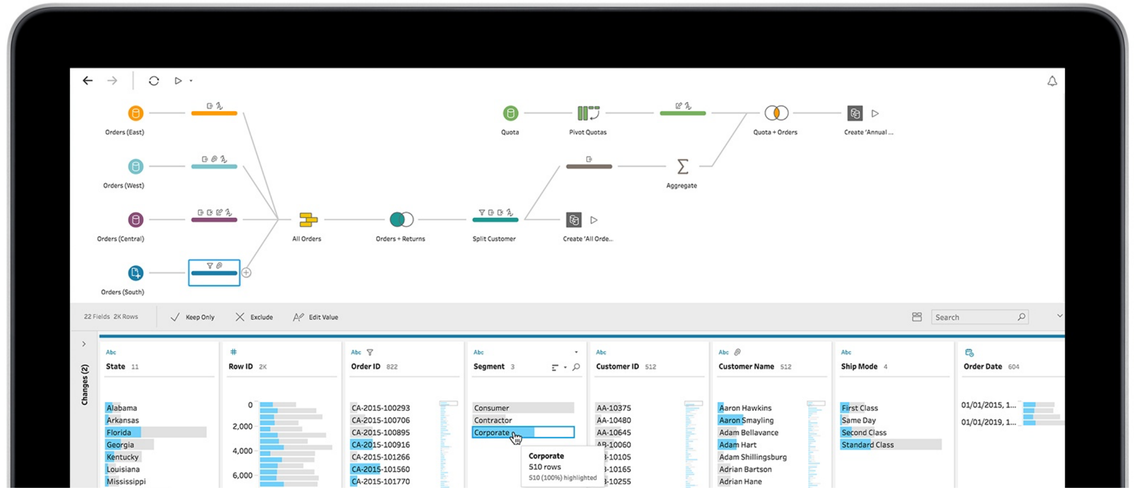


Tableau Prep Builder is an **in-memory data processing tool** with multiple data connectors. It helps us to prepare the data and export the output as CSV or Tableau Data File.

Approach



- Focus on all-type of users (not only developers)
- Multiplatform
- Enterprise-ready, proprietary and based on licenses
- Based on data flows
- Easy to create, maintain and configure
- Well integrate in the Tableau Ecosystem

Installation

Can I install Tableau Prep Builder?

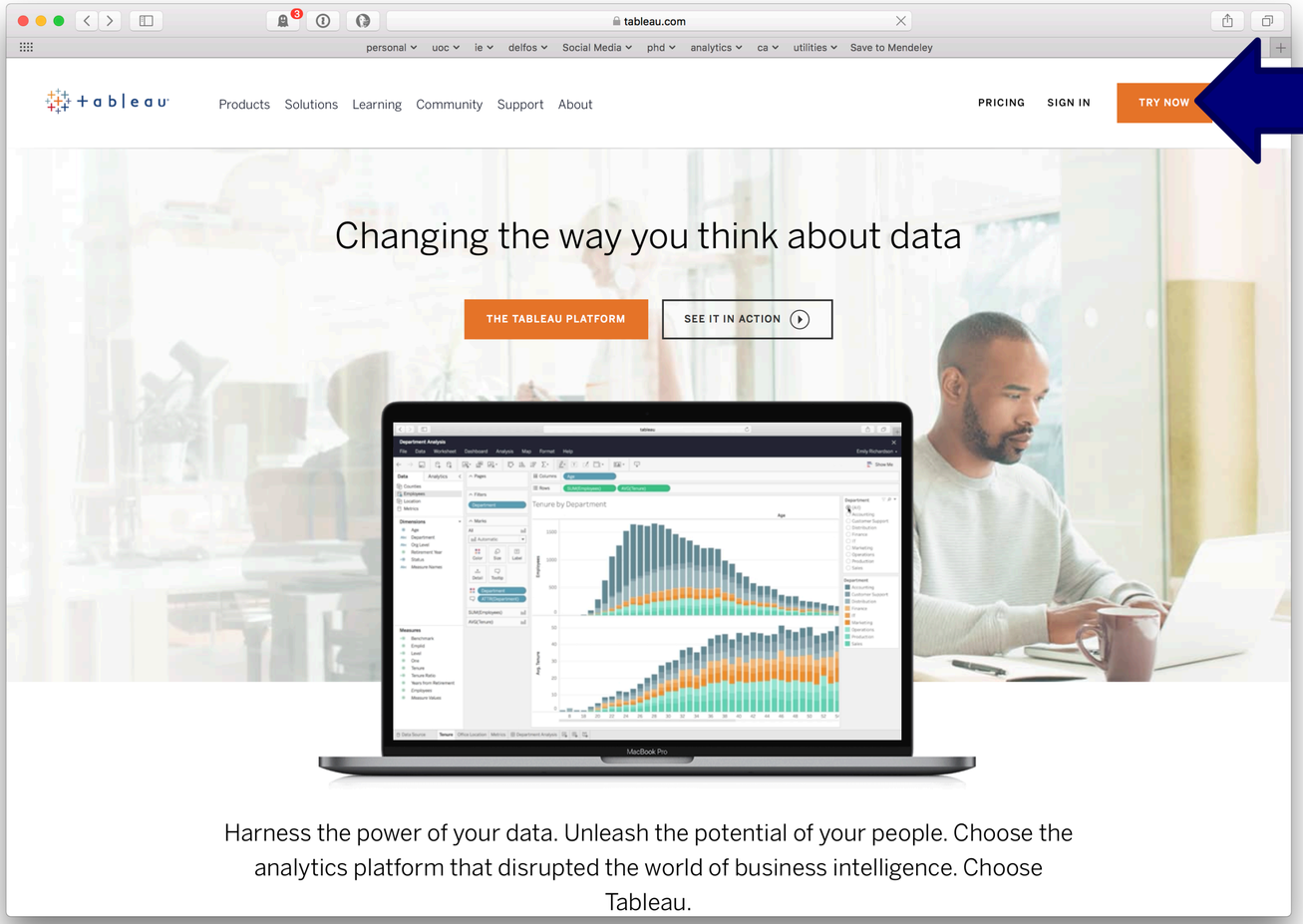
Windows

- Windows 7 or newer (64 bit)
- Intel Core i3 or AMD Ryzen 3 Pro or faster
- 4 GB memory
- 2 GM minimum free disk space

Mac

- Mac Os 10.11
- Intel Core i3 or faster
- 4 GB memory
- 2 GB minimum free disk space

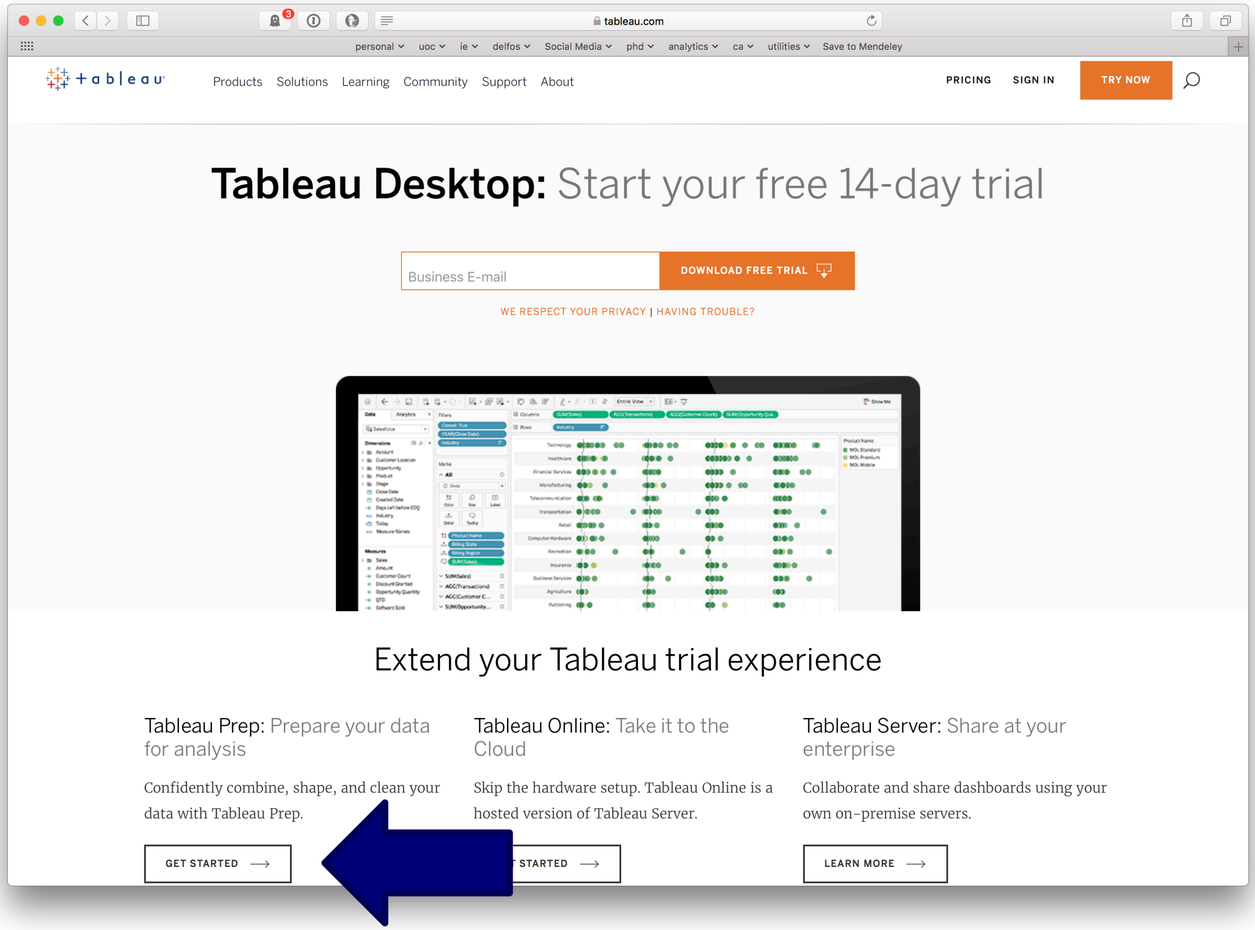
How to install Tableau Prep Builder? (I)



- Go to: https://www.tableau.com
- Click ***Try Me***

Note: At IE HST we use the provided landing page (check campus). Page may look different.

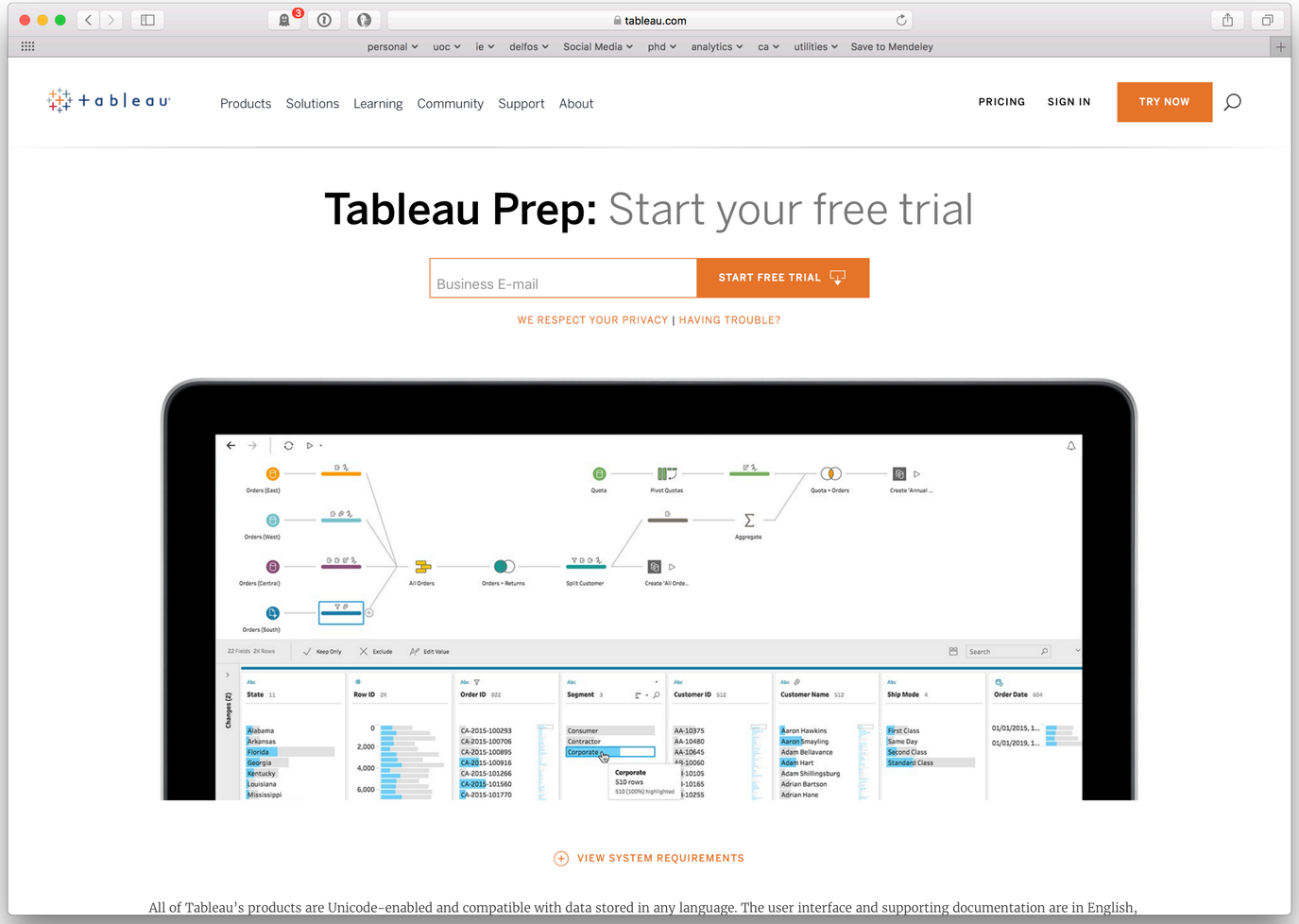
How to install Tableau Prep Builder? (II)



- W are in: https://www.tableau.com/products/trial
- Click ***GET Started***

Note: Page may look different.

How to install Tableau Prep Builder? (III)



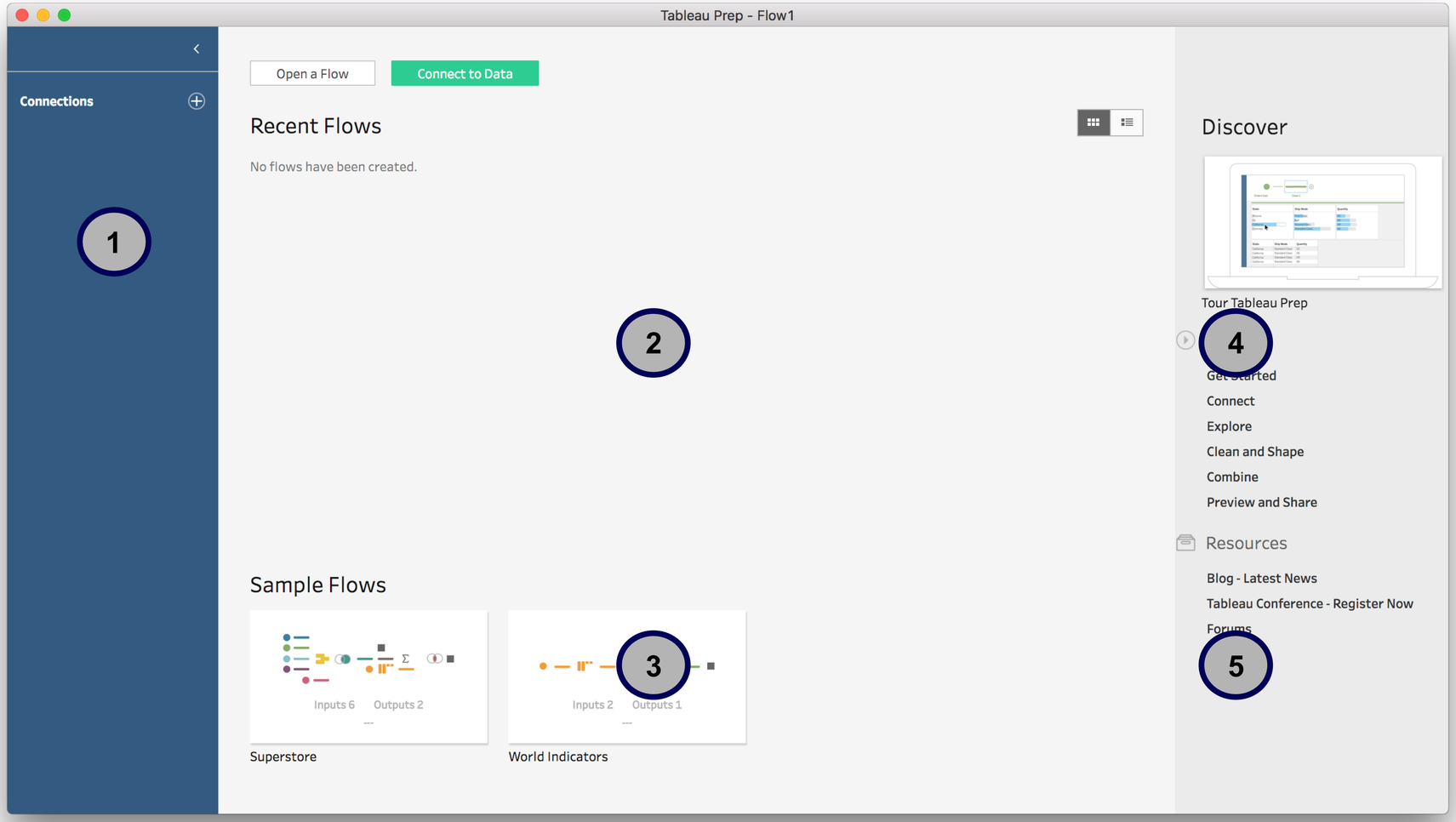
- Complete your email and **START FREE TRIAL**
- Download will start (speed will vary based on your bandwidth)
- Trail last 14 days
- Add your IE HST license

Note: Page may look different.

(Working with)
Flows

Tableau Prep Builder Interface

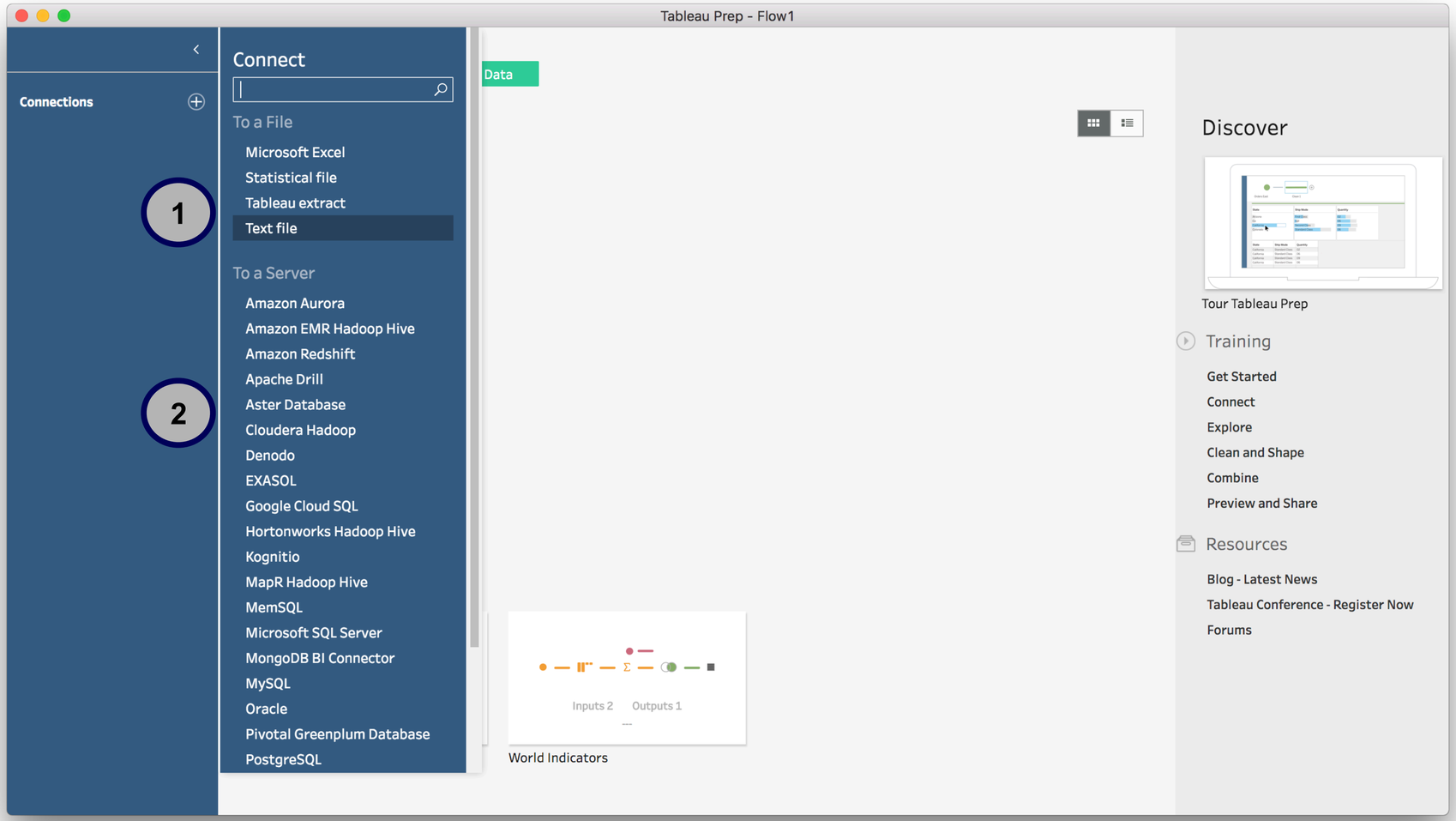
When we open the program, this is the first screen…



1. **Connections**
2. **Previous Flows**
3. **Sample Flows**
4. **Training**
5. **Resources & Updates**

Data Connectors

There are multiple connectors:



1 Files

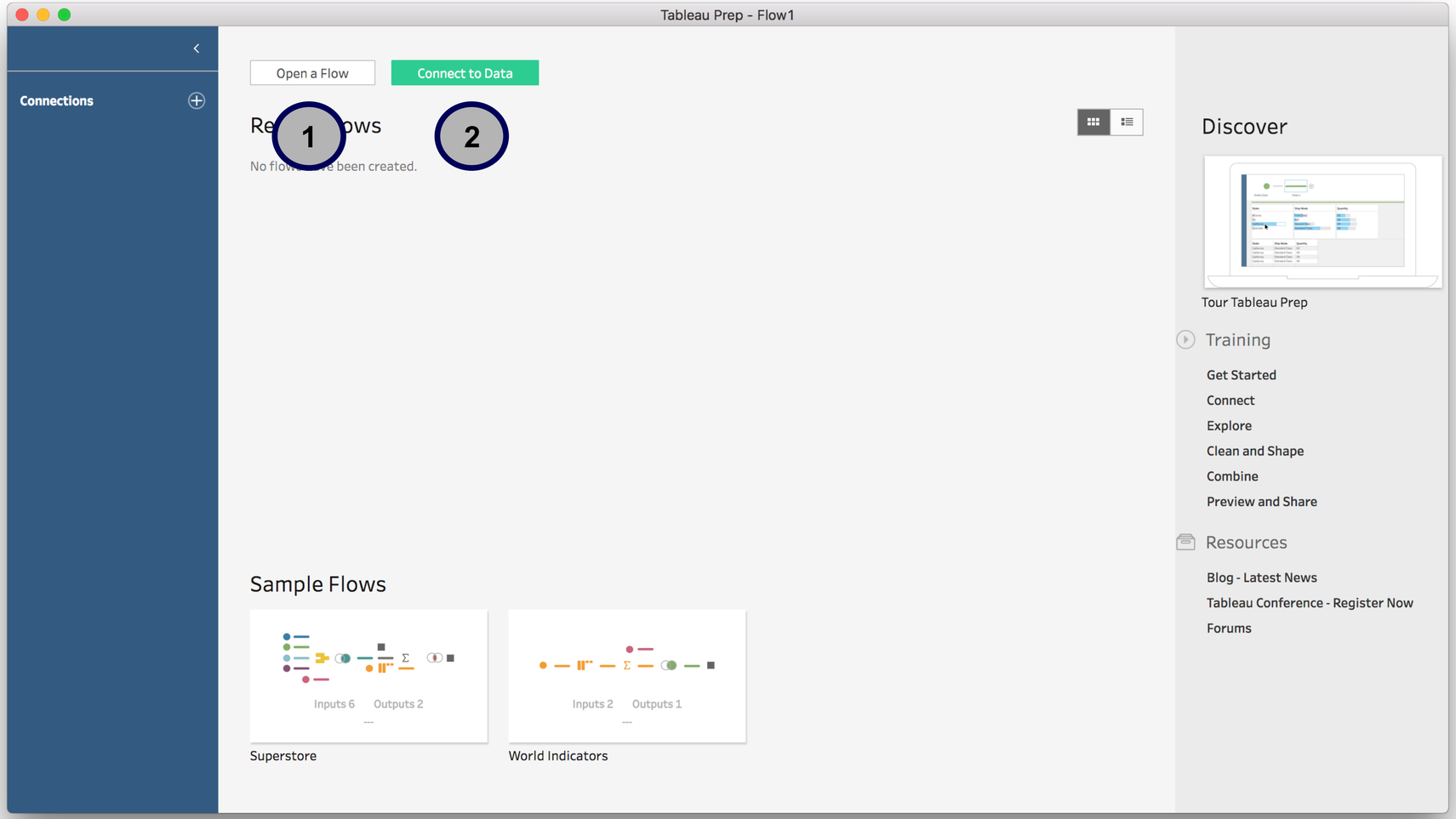
Including excel, text, tableau files and others

2 Servers & Databases

Including all type of databases, big data systems and enterprise and commercial applications

How to work with Tableau Prep Builder

Two options:



1 **Flows**

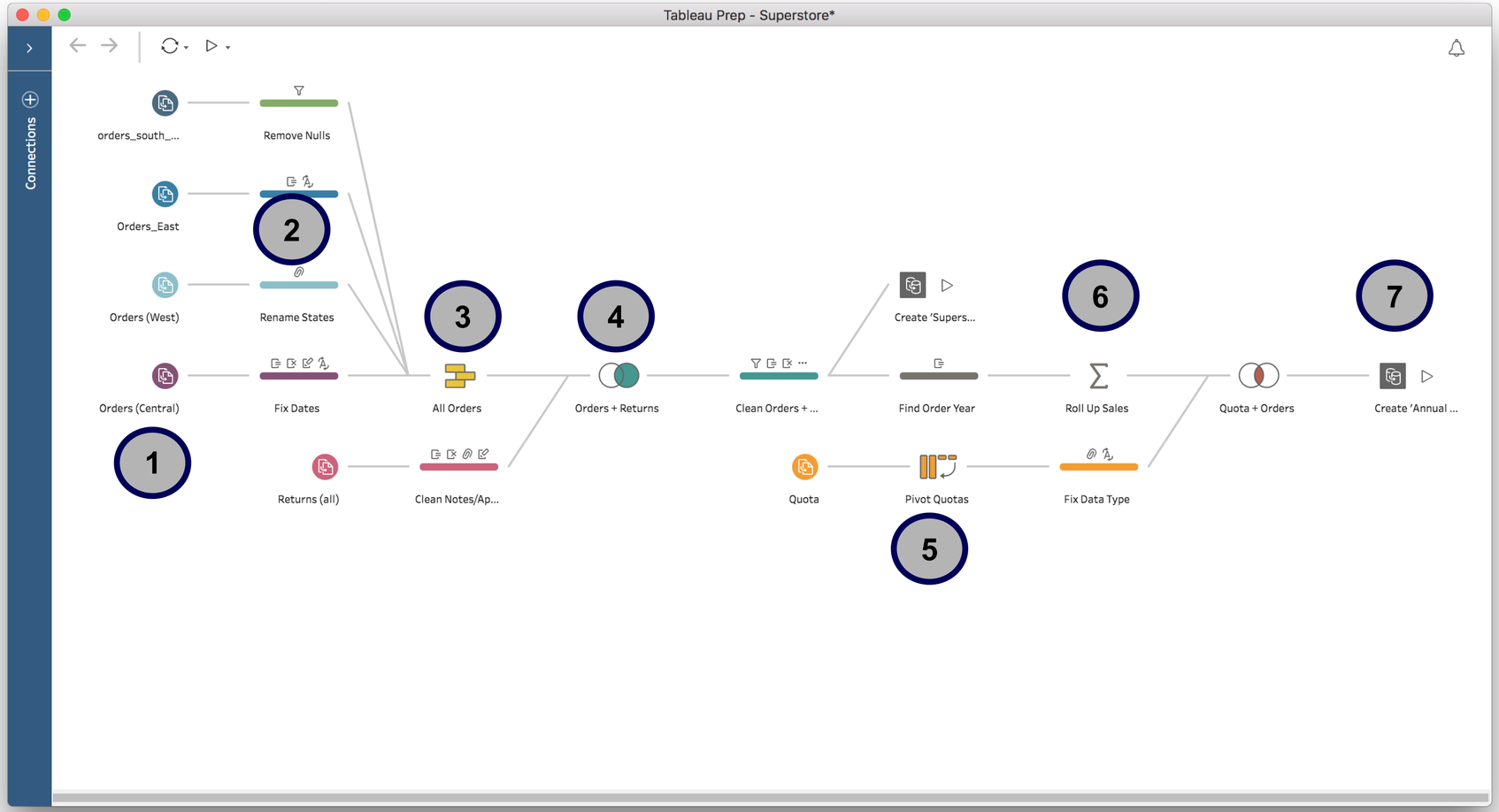
We open a previous flow (to execute or modify it)

2 **Data**

We open a data source (to create a new flow)

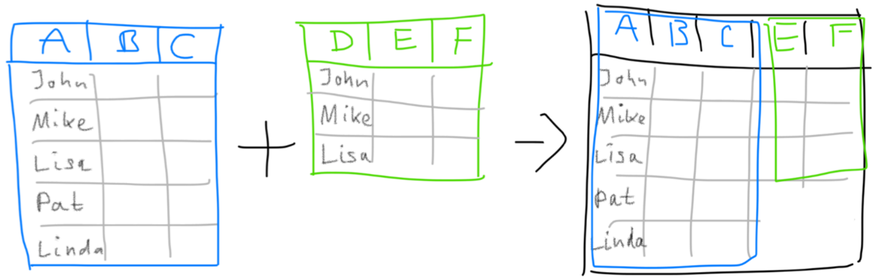
What is a Flow?

A collection of steps that we apply to a data source. This is an ETL process (Extract, Transform and Load).
We have several options:

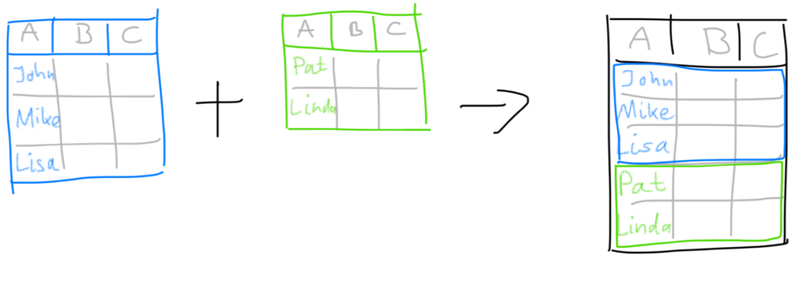


1. **Access data**
2. **Clean data**
3. **Union data** (based on the same fields)
4. **Join data** (all, intersection,…)
5. **Pivot data** (create a pivot table)
6. **Aggregate data** (based on a metric or attribute)
7. **Export**

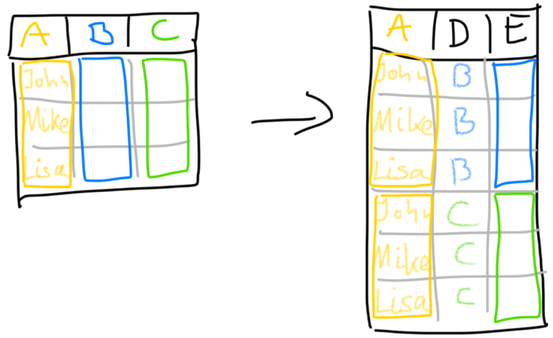
Relevant steps in a flow



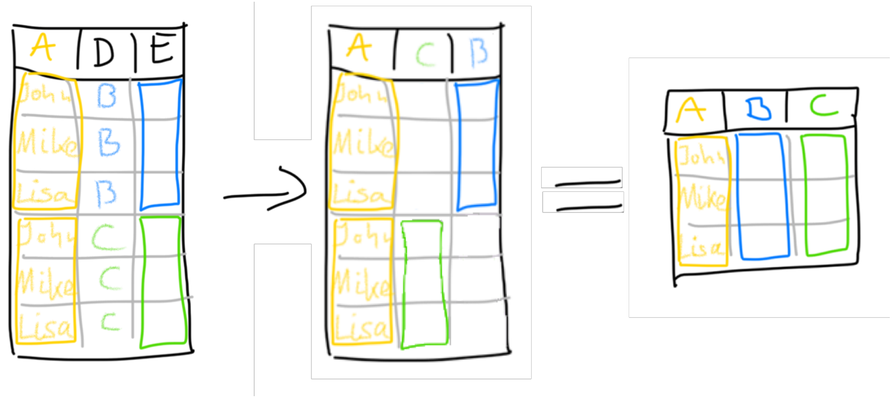
Join



Union



Pivot



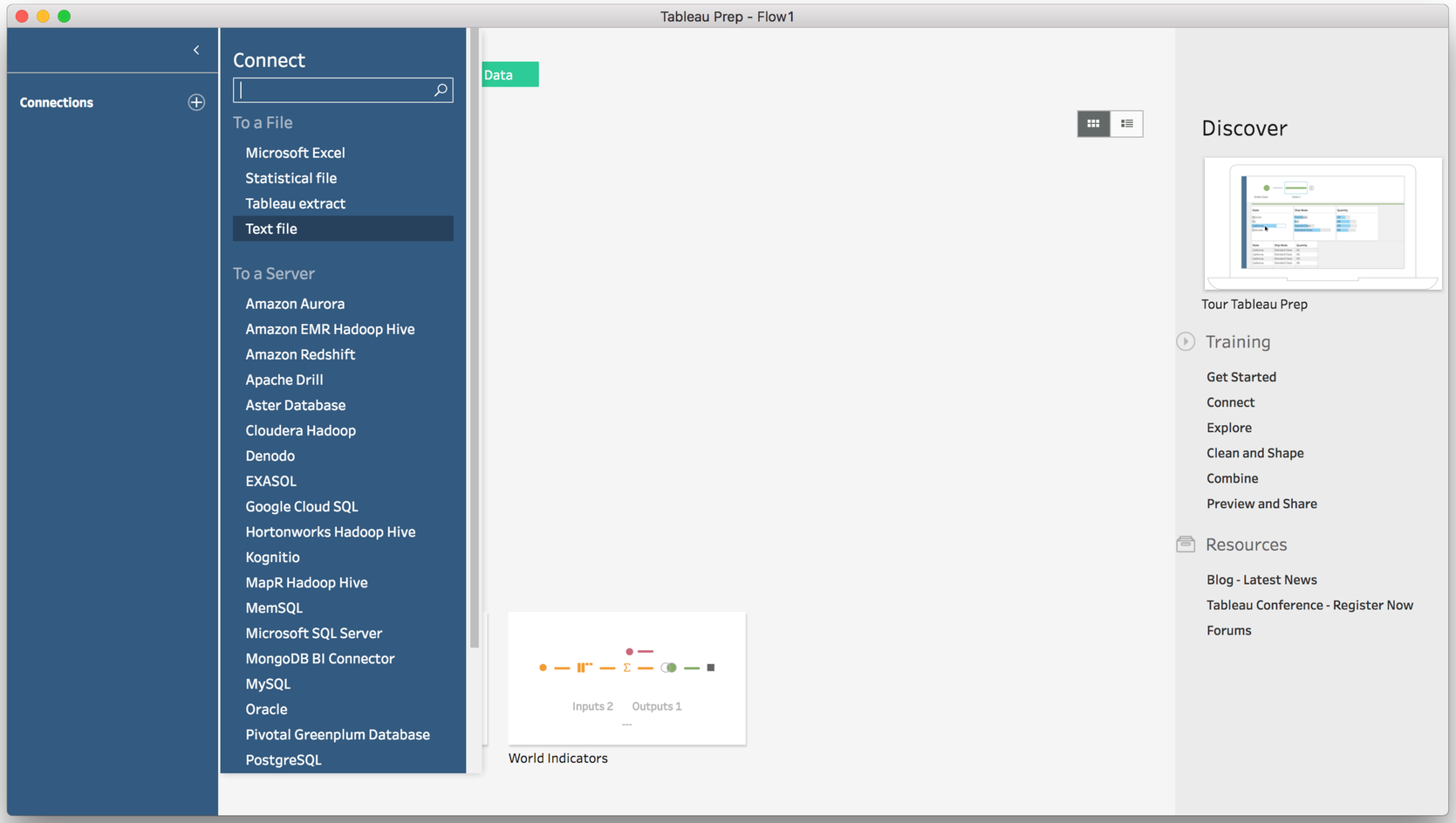
Unpivot

- When we have several data sources, the main options to combine them are: join, union, pivot and unpivot.
- The others steps are related to data cleaning and preparation.

Example

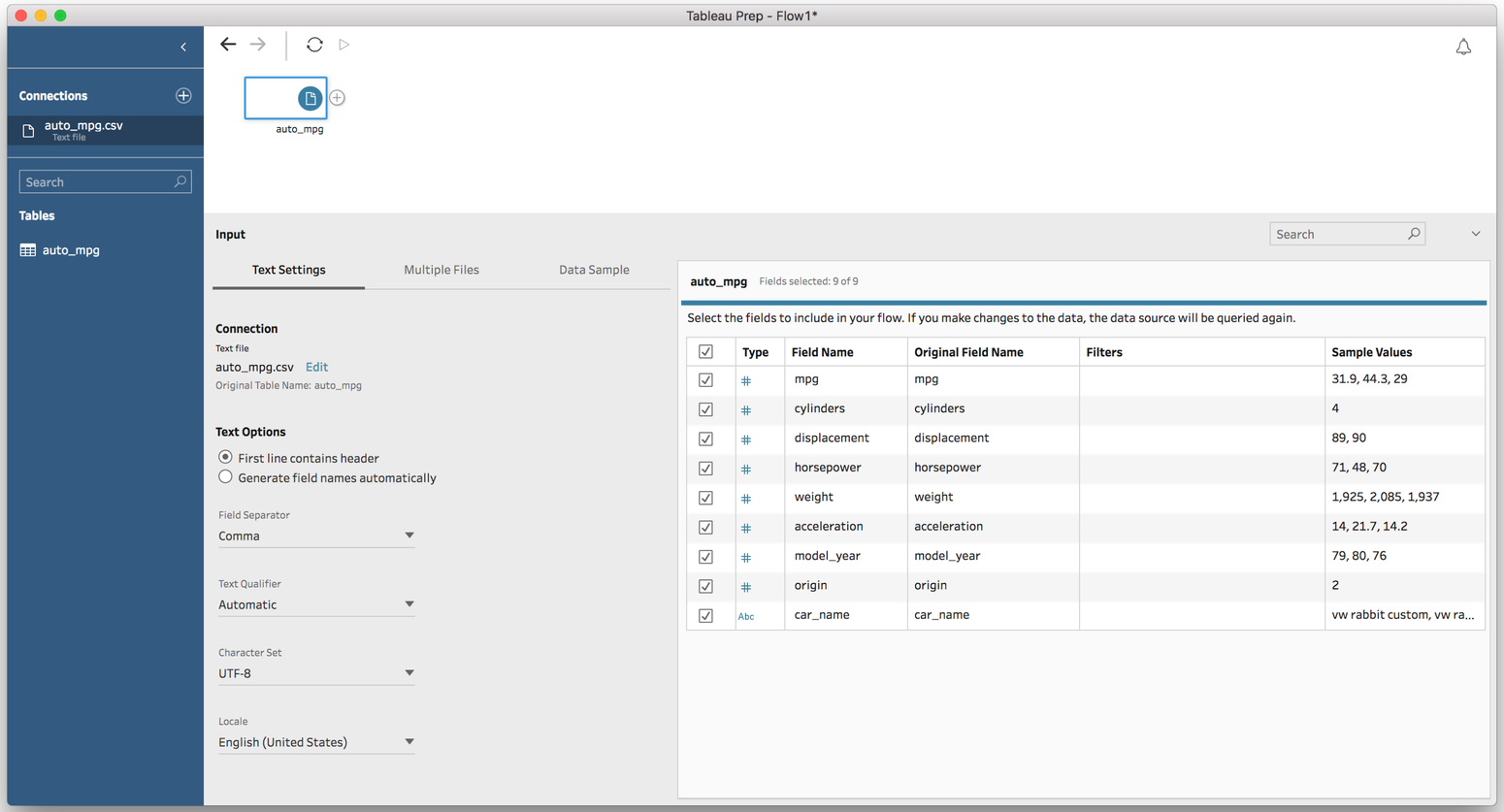
(we recommend to check this link as well:
https://onlinehelp.tableau.com/current/prep/en-us/prep_welcome.htm)

Acces data (I)



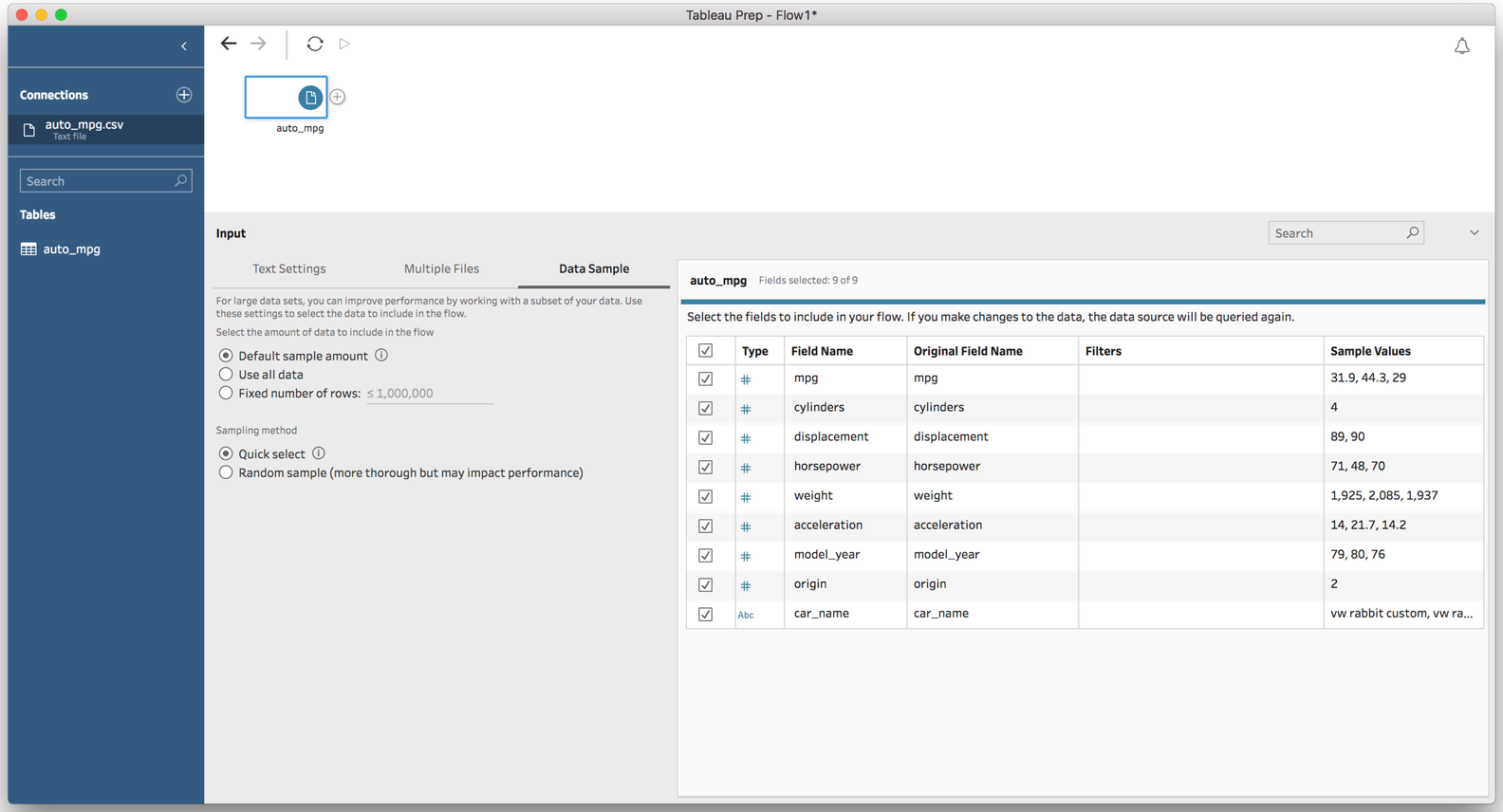
- We will prepare the dataset: **auto_mpg.csv**
- Click on **Connections**
- Select **text file**

Acces data (II)



- For each connection we can change the options.
- In this case: *text settings*, *multiple files* y *data sample*.
- For *text settings*: format, separation, header, etc.

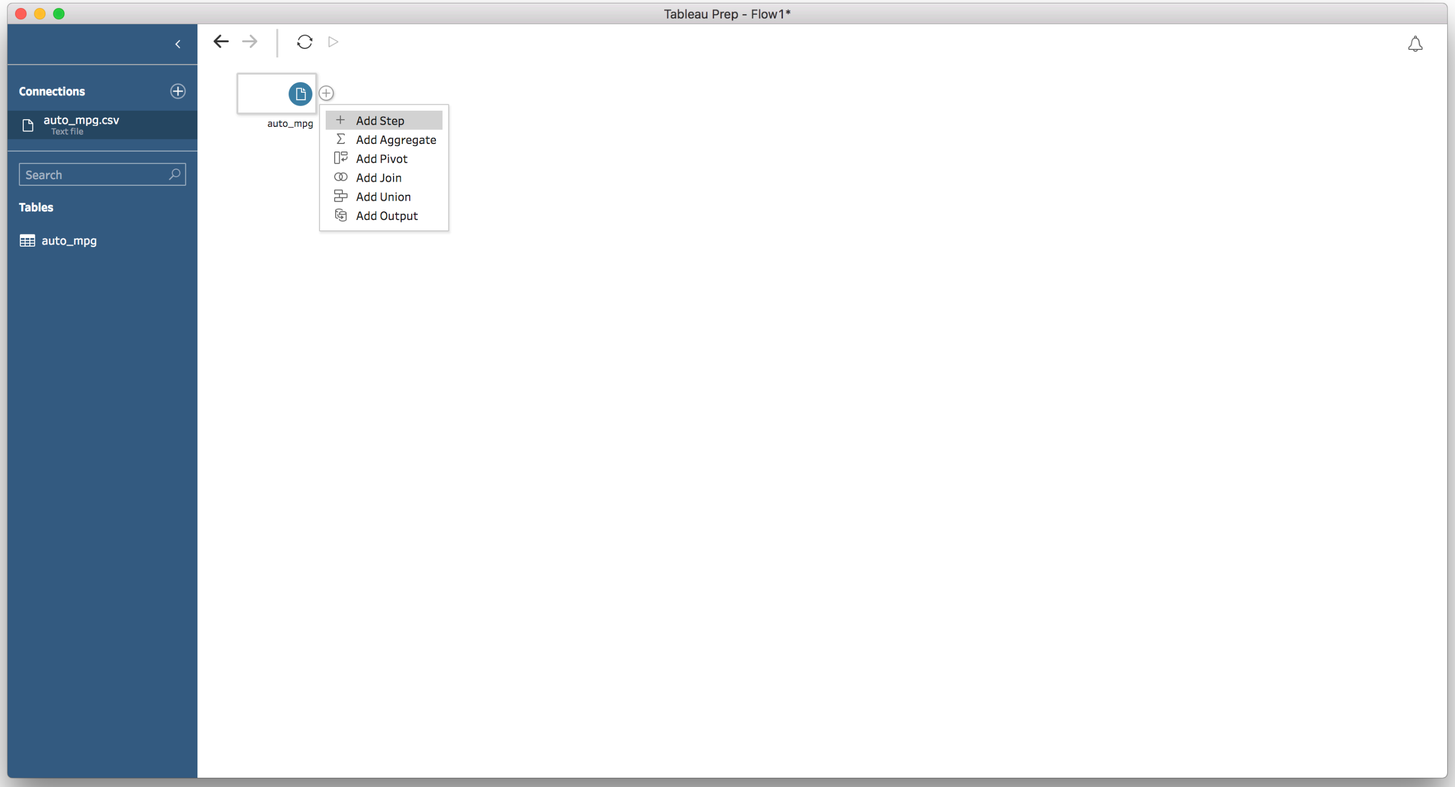
Acces data (III)



- If the file is small, we can use all the data (this case).
- If not use a sample (default or by number of rows)

Data Preparation (I)

After loading the dataset, we have our step and we need to add the next one. In this step, we will start preparing the data. Click (+) and we will display a menu with all the options.

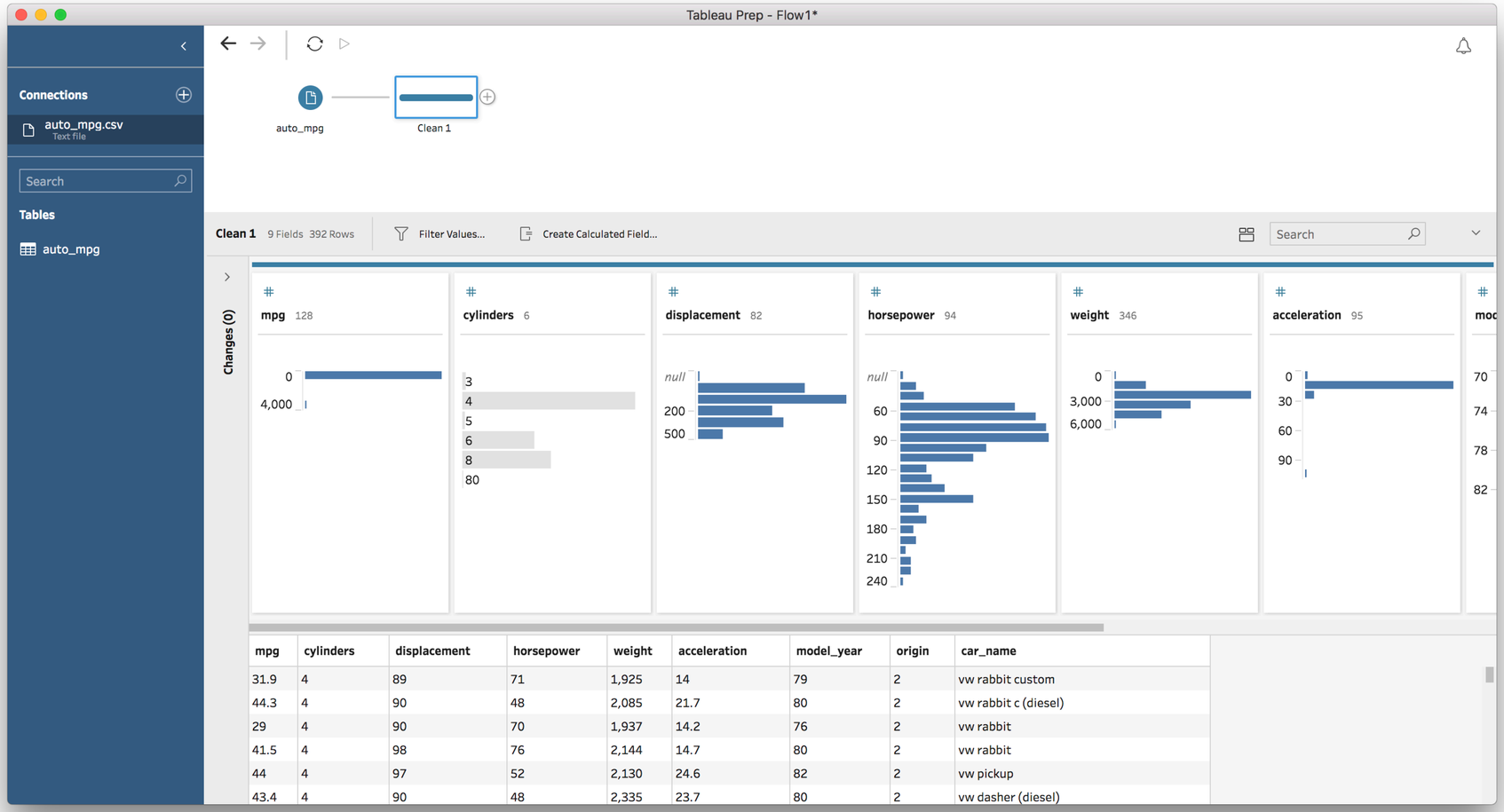


- We have six steps
- ***Step:*** to apply changes and cleasing functions to our data.
- ***Aggregate*, *pivot*, *join* y *union***: to combine the flow of data
- **Output**: to export the flow to a new data set.

Check this link for more videos about all the steps in Tableau Prep Builder: https://www.tableau.com/learn/training#prep

Data Preparation (II)

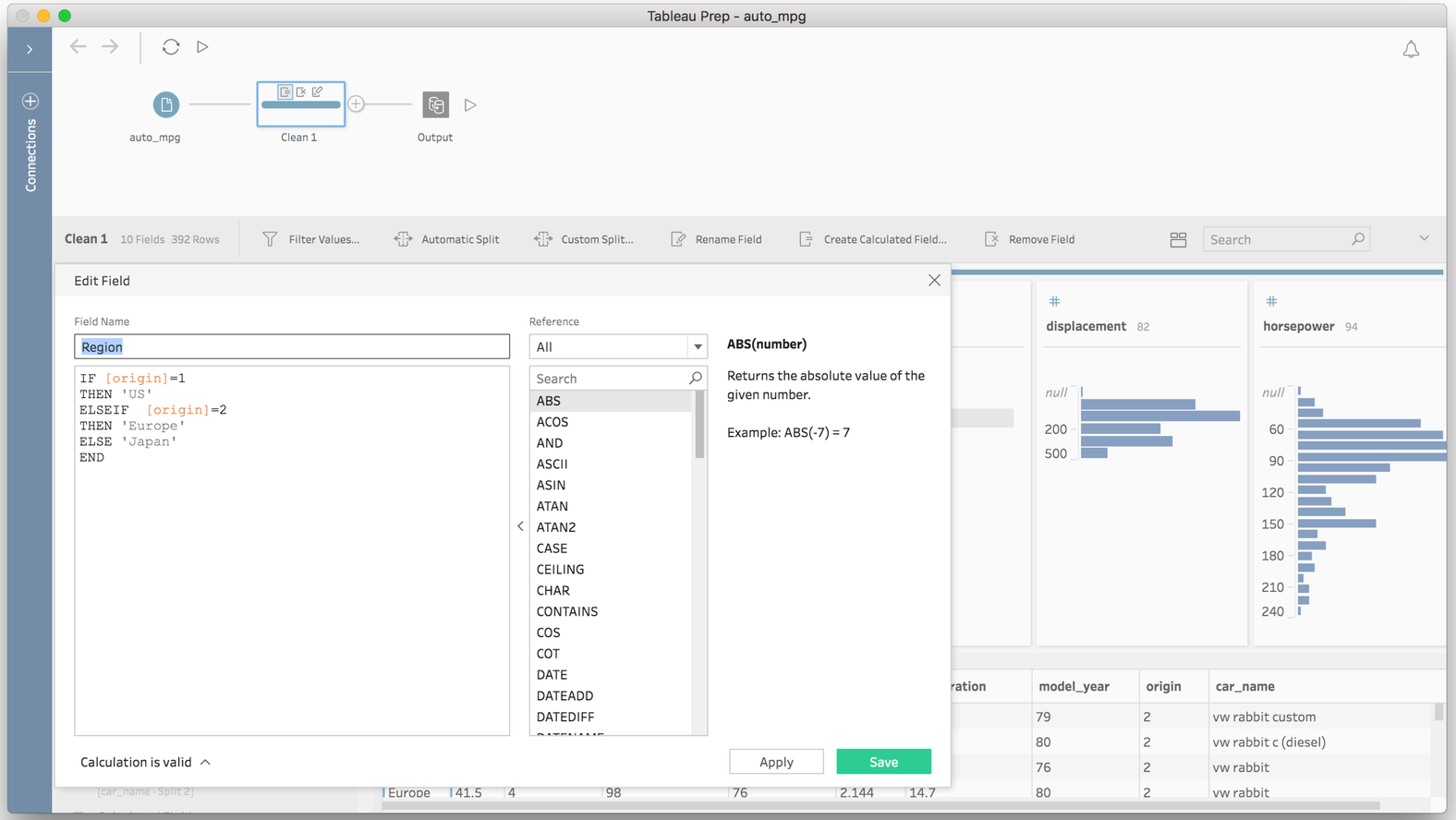
As starting point, we will prepare the data in auto_mpg.csv



- When selecting the first *step*, we will have the distribution per field. This can help us to spot error to solve and transformations to apply.
- One thing that we can do is create calculated fields.

Data Preparation (III)

Using *calculated field*, we can create new fields in our data set.

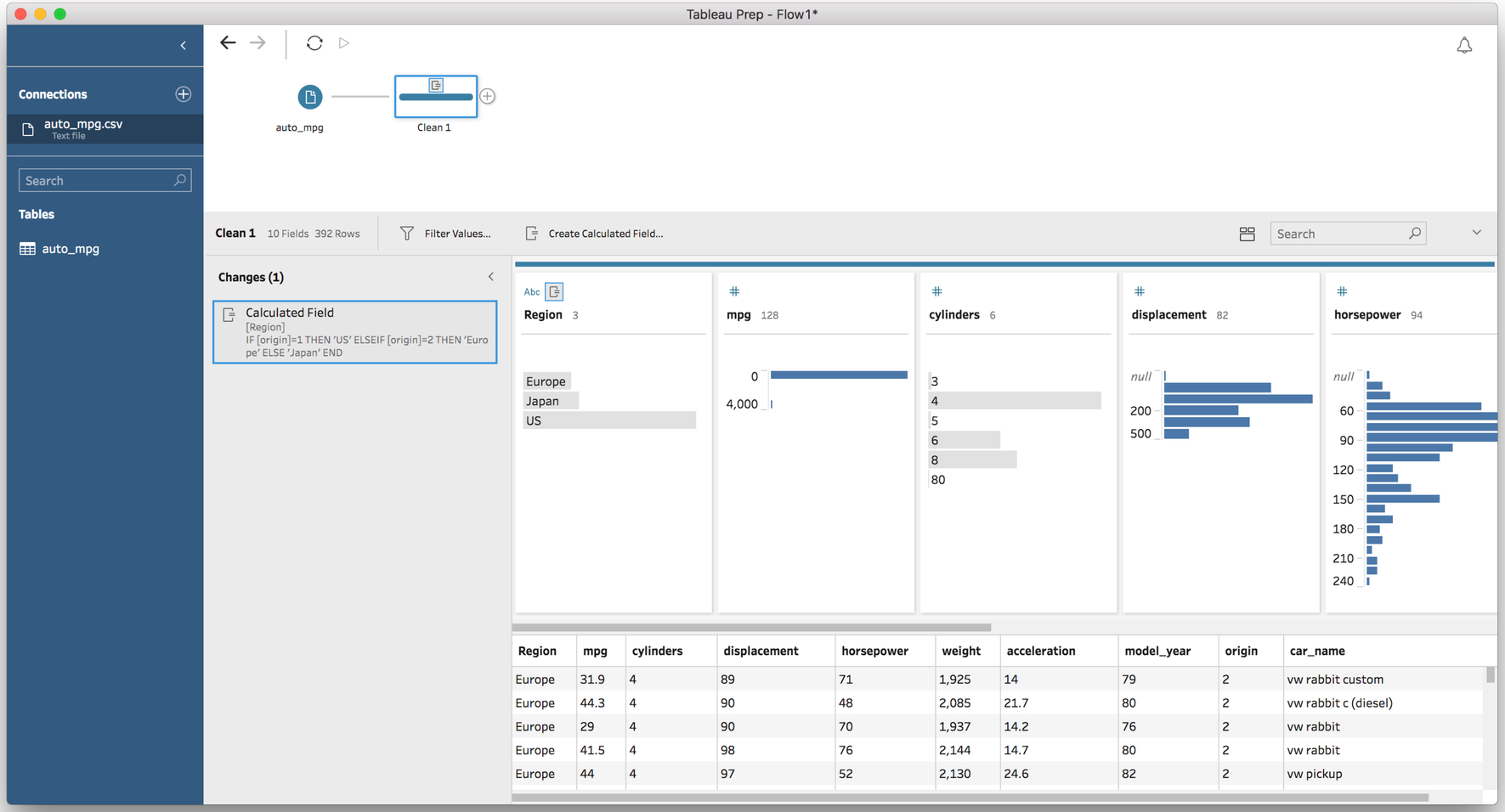


- Let's create a new field called **region** based on **origin**.

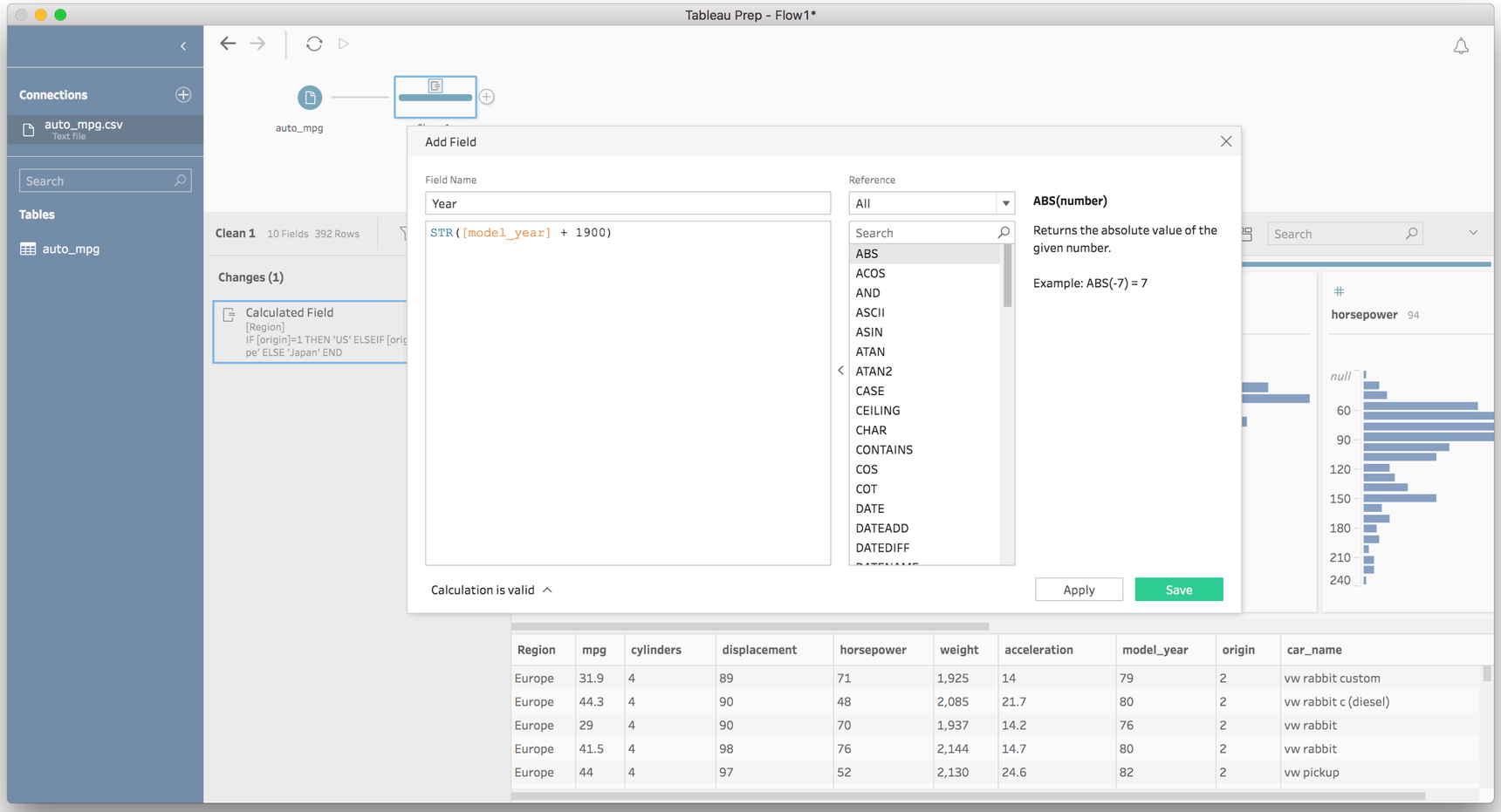
```
IF [origin]=1
THEN 'US'
ELSEIF  [origin]=2
THEN 'Europe'
ELSE 'Japan'
END
```

Data Preparation (IV)

Every time that will apply a transformation, we will find it in *changes* as a sequence:



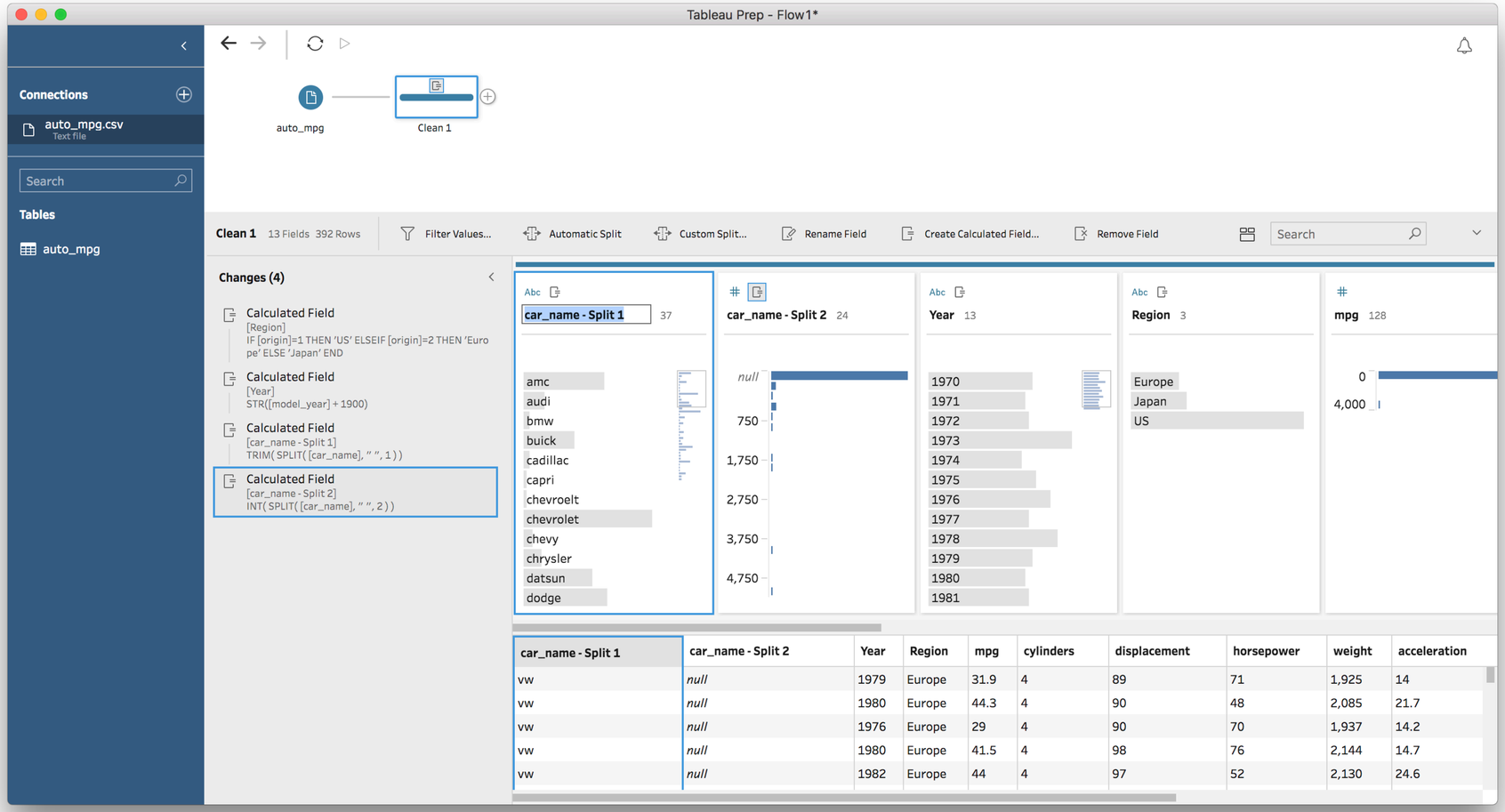
Data Preparation (V)



- We can create **”Year”** from "Model Year" using ***calculated field***:

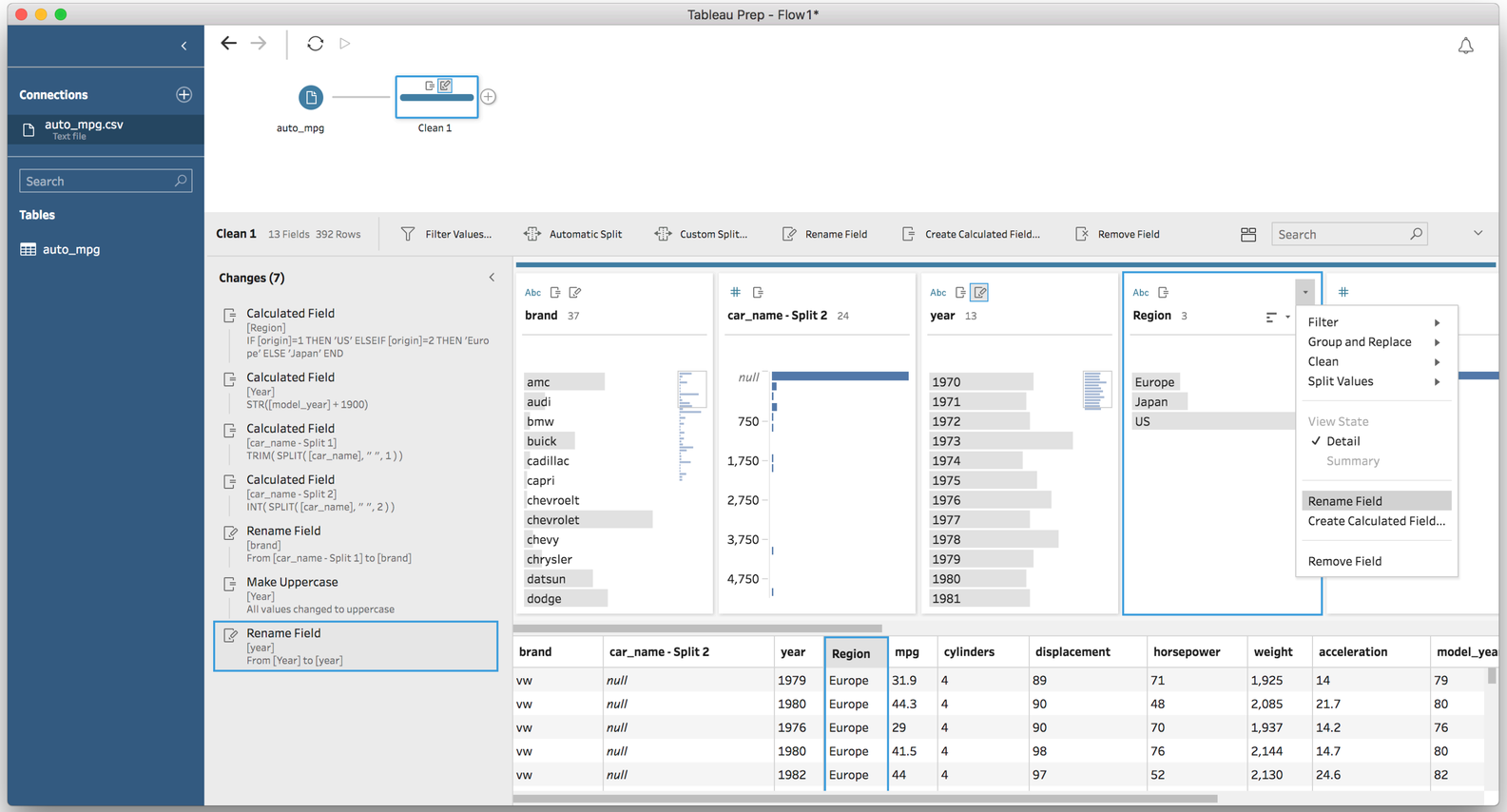
STR([Model Year] + 1900)

Data Preparation (VI)



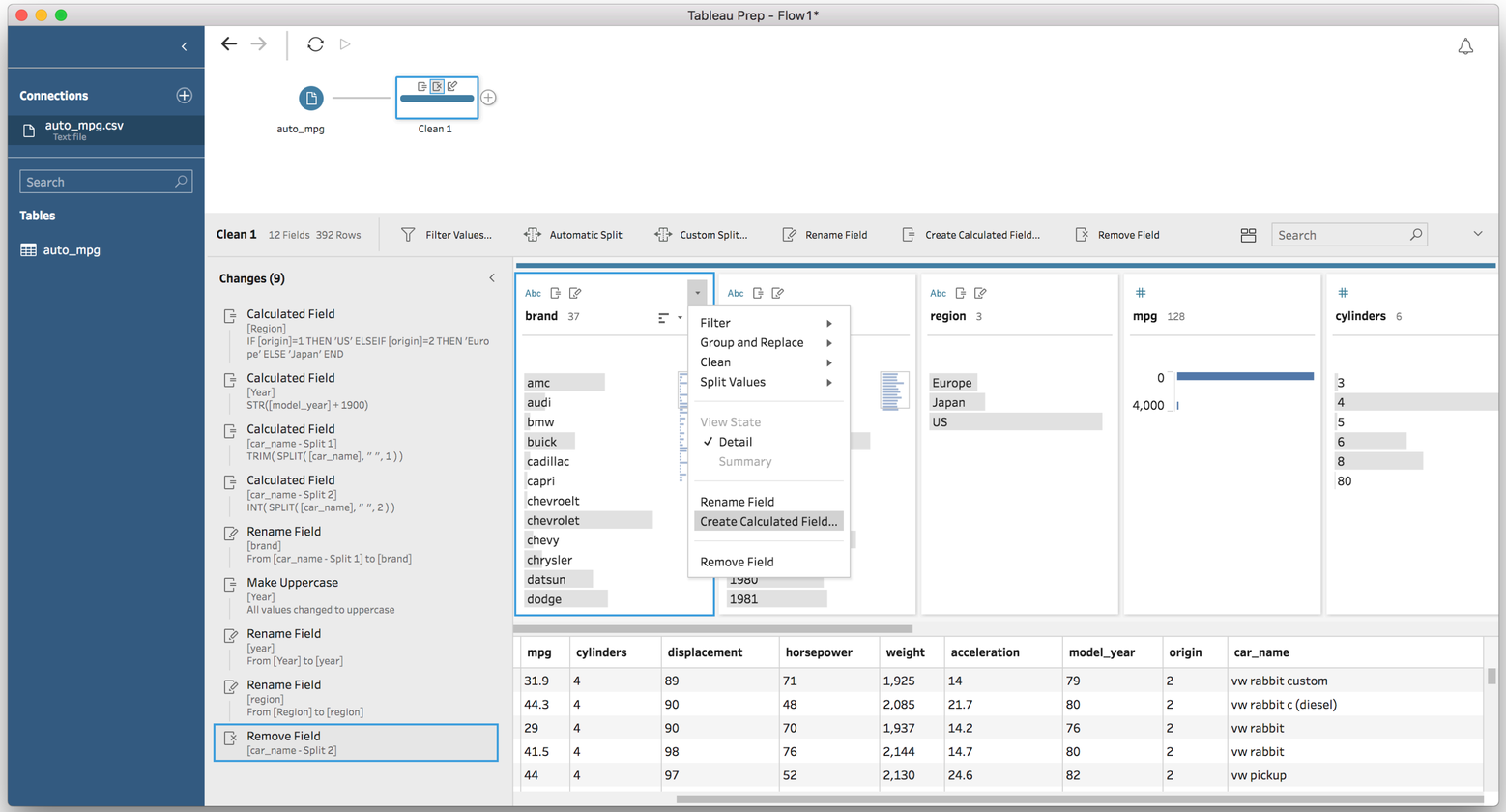
- We can create **“Brand”** applying *split* to "Car Name”.
- Two columns are generated: split 1 and 2.
- We need the first one only.
- We will delete the second one with *Remove Field*

Data Preparation (VII)



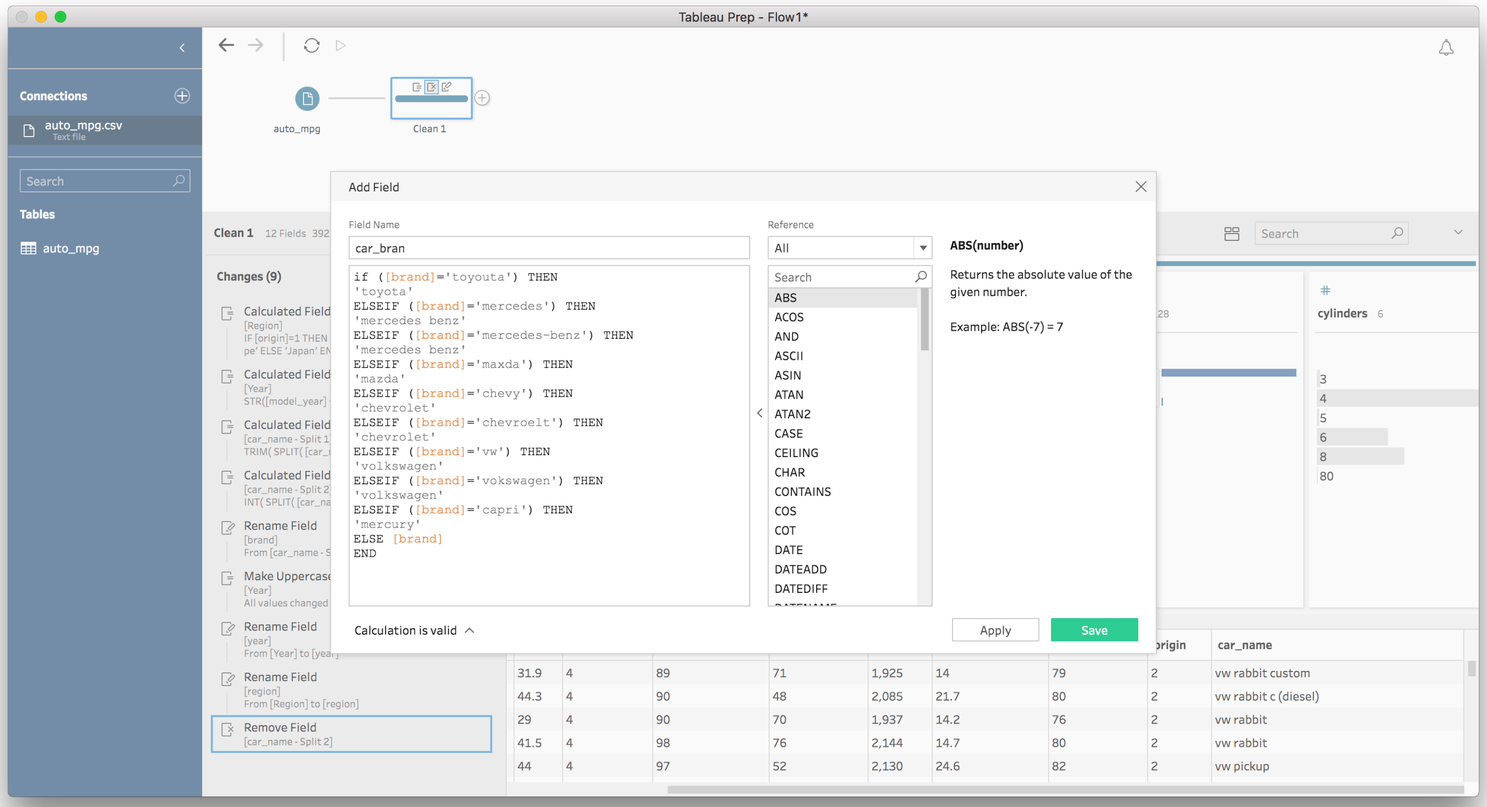
- If we make a mistake with a field name we can change it (using *rename field*).
- In every menu we can find several options that can help us.

Data Preparation (VIII)



- After deleting split 2 borrar split 2, we have detected that there are some errors in the brand names. We must correct them.
- We will create another calculated field.

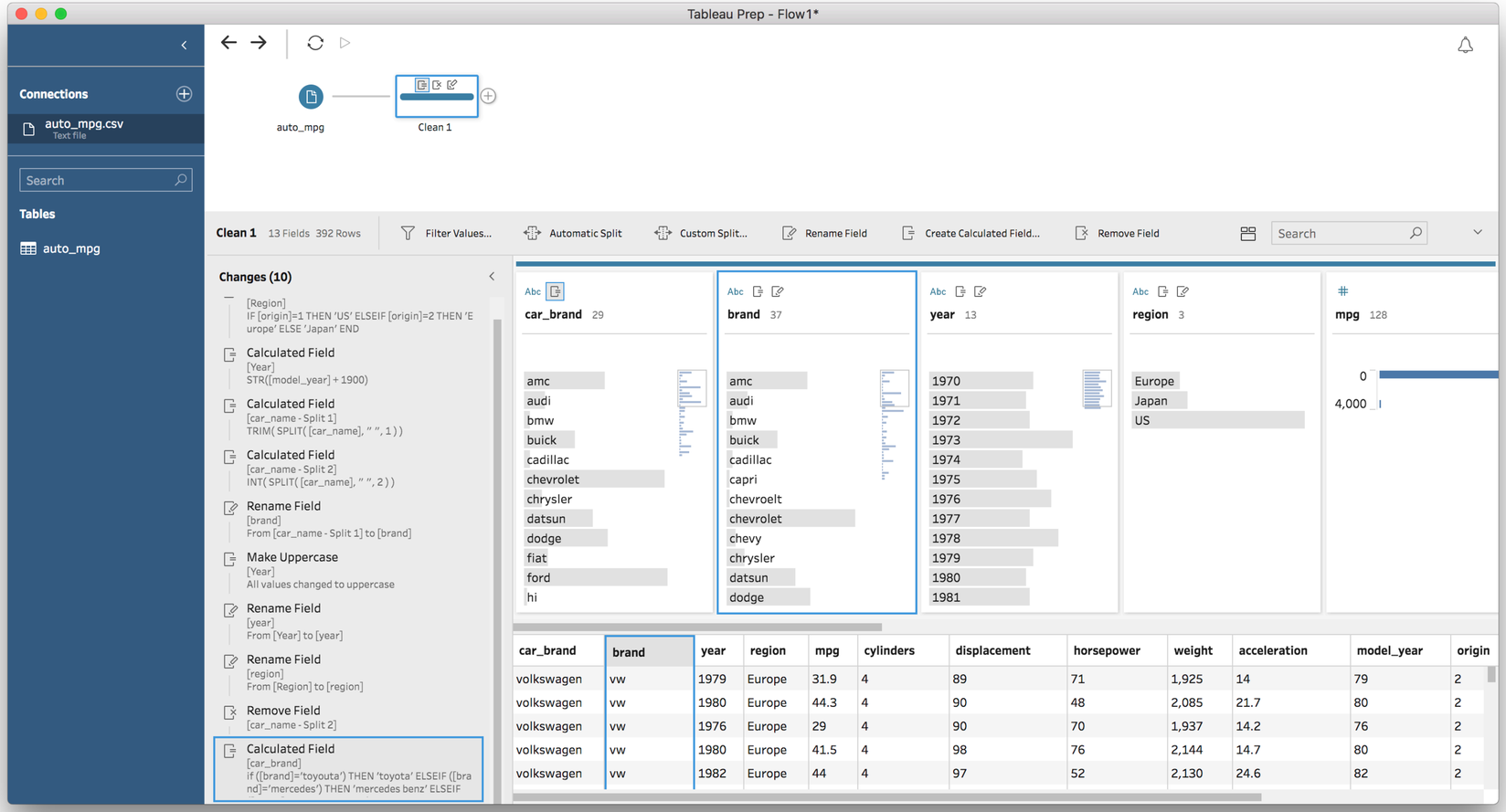
Data Preparation (IX)



- We will create **"Card Brand"** using *calculated field*:

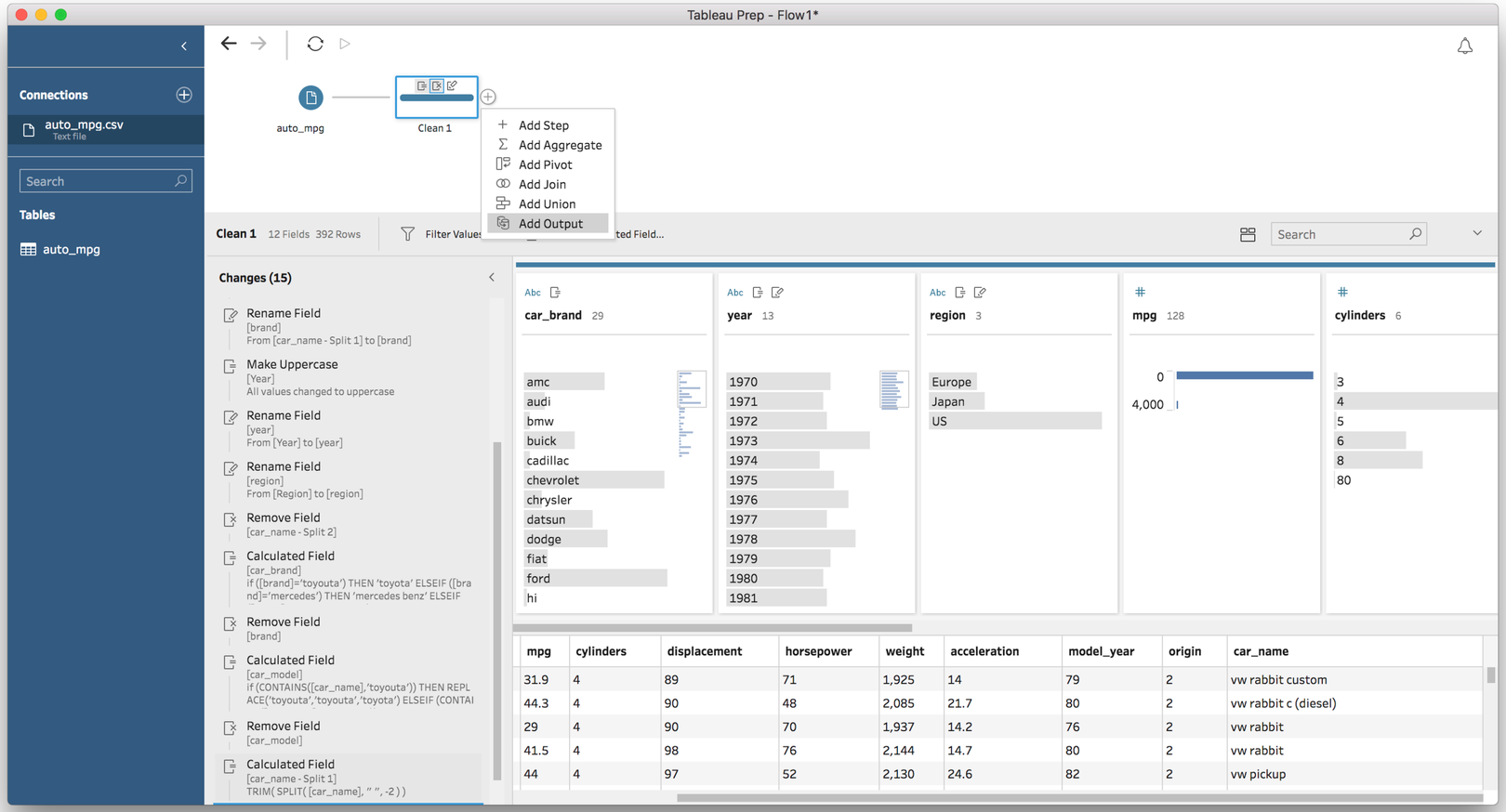
```
if ([Brand]='toyouta') THEN
'toyota'
ELSEIF ([Brand]='mercedes') THEN
'mercedes benz'
ELSEIF ([Brand]='mercedes-benz') THEN
'mercedes benz'
ELSEIF ([Brand]='maxda') THEN
'mazda'
ELSEIF ([Brand]='chevy') THEN
'chevrolet'
ELSEIF ([Brand]='chevroelt') THEN
'chevrolet'
ELSEIF ([Brand]='vw') THEN
'volkswagen'
ELSEIF ([Brand]='vokswagen') THEN
'volkswagen'
ELSEIF ([Brand]='capri') THEN
'mercury'
ELSE [Brand]
END
```

Data Preparation (X)



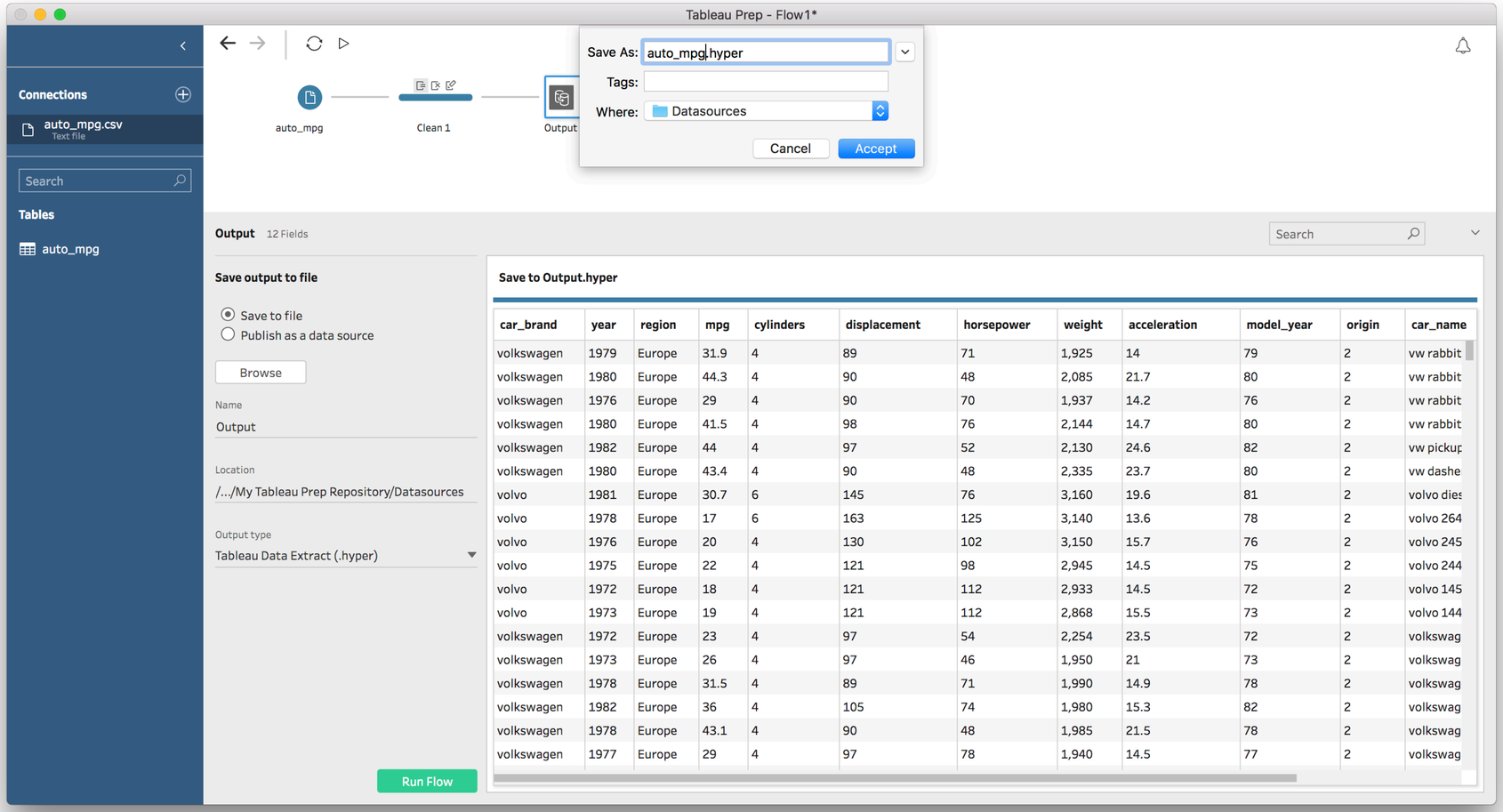
- After creating **car_brand,** we can delete **brand** (as you can imagine only if we don't need this field anymore)**.**

Data Output (I)



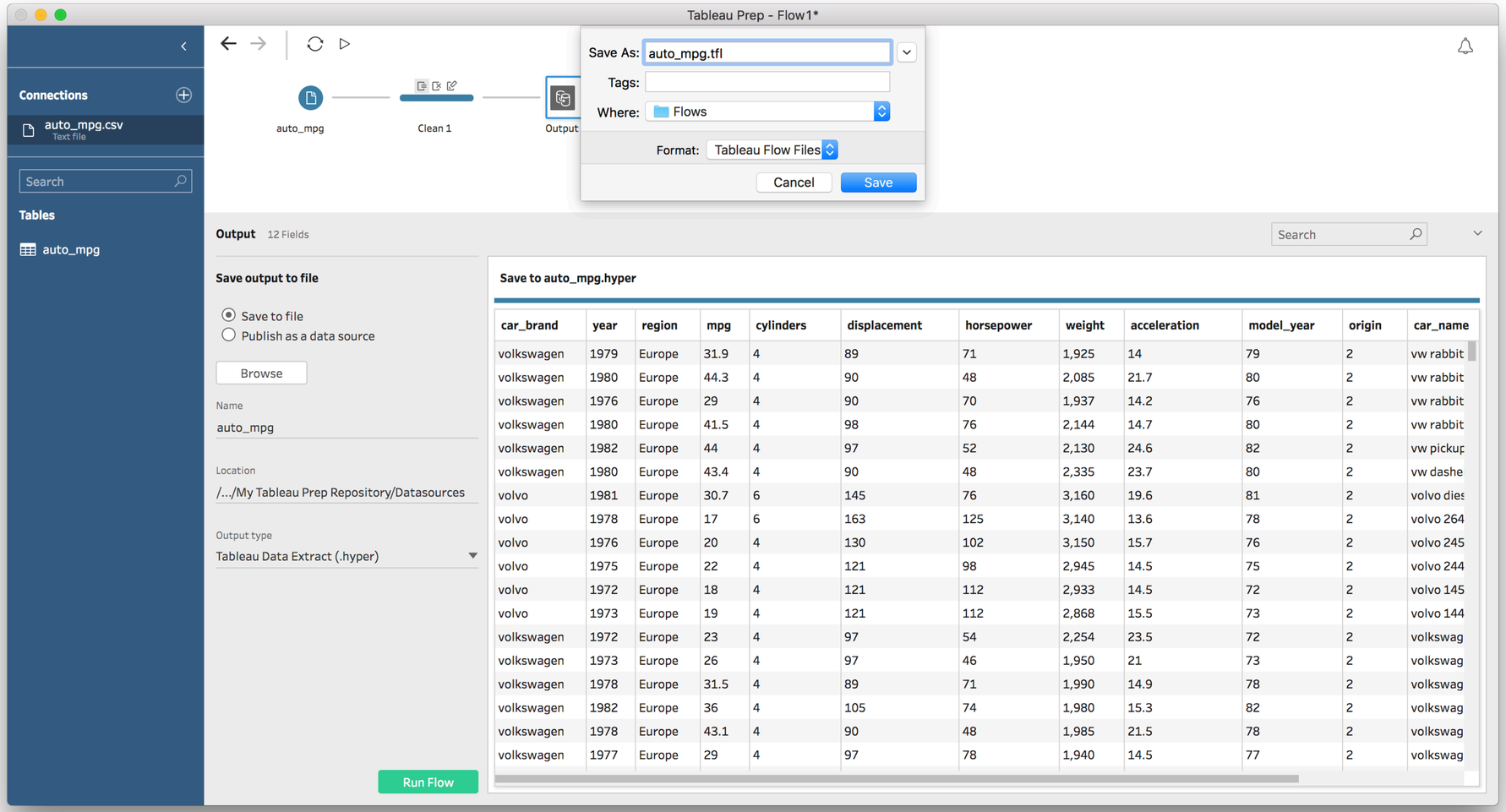
- After the data preparation, we must create the output file and flow.
- That means a new file that contains all the changes (and we will use during our analysis).

Data Output (II)



- We must select the output format (recommended hyper).
- Remember to choose a proper name and a folder.

Data Output (III)



car_brand	year	region	mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	car_name
volkswagen	1979	Europe	31.9	4	89	71	1,925	14	79	2	vw rabbit
volkswagen	1980	Europe	44.3	4	90	48	2,085	21.7	80	2	vw rabbit
volkswagen	1976	Europe	29	4	90	70	1,937	14.2	76	2	vw rabbit
volkswagen	1980	Europe	41.5	4	98	76	2,144	14.7	80	2	vw rabbit
volkswagen	1982	Europe	44	4	97	52	2,130	24.6	82	2	vw pickup
volkswagen	1980	Europe	43.4	4	90	48	2,335	23.7	80	2	vw dashe
volvo	1981	Europe	30.7	6	145	76	3,160	19.6	81	2	volvo dies
volvo	1978	Europe	17	6	163	125	3,140	13.6	78	2	volvo 264
volvo	1976	Europe	20	4	130	102	3,150	15.7	76	2	volvo 245
volvo	1975	Europe	22	4	121	98	2,945	14.5	75	2	volvo 244
volvo	1972	Europe	18	4	121	112	2,933	14.5	72	2	volvo 145
volvo	1973	Europe	19	4	121	112	2,868	15.5	73	2	volvo 144
volkswagen	1972	Europe	23	4	97	54	2,254	23.5	72	2	volkswag
volkswagen	1973	Europe	26	4	97	46	1,950	21	73	2	volkswag
volkswagen	1978	Europe	31.5	4	89	71	1,990	14.9	78	2	volkswag
volkswagen	1982	Europe	36	4	105	74	1,980	15.3	82	2	volkswag
volkswagen	1978	Europe	43.1	4	90	48	1,985	21.5	78	2	volkswag
volkswagen	1977	Europe	29	4	97	78	1,940	14.5	77	2	volkswag

- We must save the flow as well.
- This way we can modify the flow or share it with the rest of the team.


Backspace
Thank You
Page Up
Home
Page Down
Insert
Shift