

## How to Get Start with Clementine

تهیه کننده:

سعید ادیب فر

پاییز ۹۶

## فهرست

بخش اول: آشنایی با نرمافزار Clementine ..... ۱

بخش دوم: کار با Stream ها ..... ۹

## توضیحات

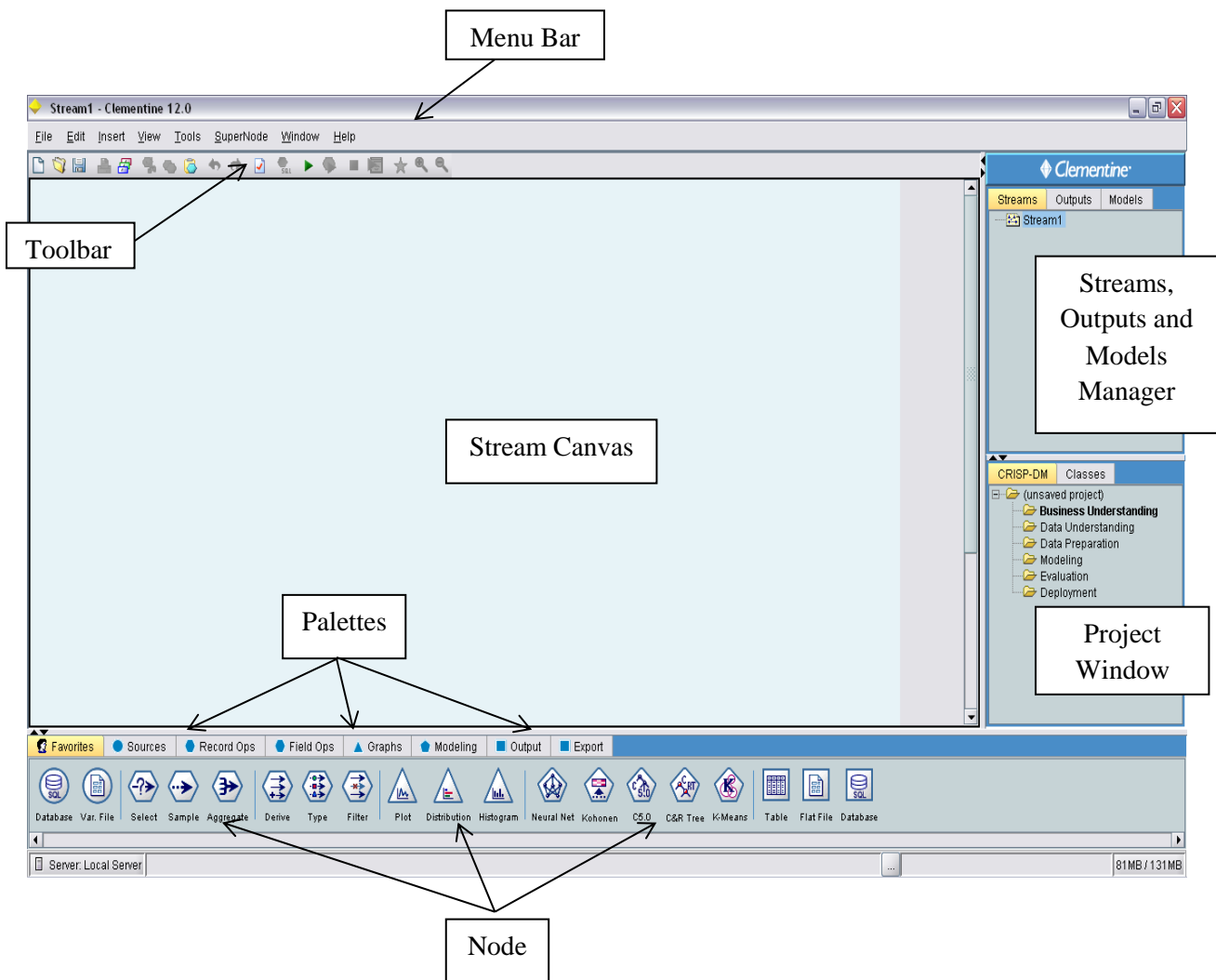
این گزارش با استفاده از Help نرم‌افزار و همچنین بهره‌گیری از اسلایدهای دکتر یقینی در آموزش نرم‌افزار Clementine، تهیه و تدوین شده است.

## بخش اول: آشنایی با نرم افزار Clementine

Clementine، قدیمی ترین و در عین حال پرتعدادترین نرم افزار داده کاوی است که شما را در ساخت مدل های پیش بینانه برای تصمیم گیری های آینده کمک می نماید.

## فضای کار نرم افزار:

محیط کار Clementine به همراه معرفی قسمت های مختلف نرم افزار در شکل زیر آمده است:

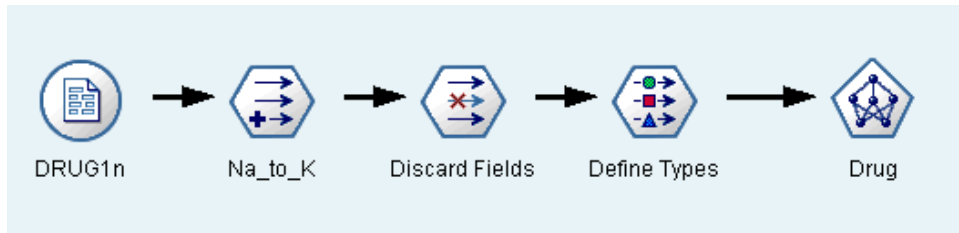


نمای ظاهری نرم افزار این اجازه را به شما می دهد تا جریان های منطقی داده هایتان را همانطور که در ذهن دارید رسم کنید. کار با Clementine یک فرآیند ۳ مرحله ای کار با داده می باشد:

۱. داده را از Clementine بخوانید.
۲. داده را از طریق یک سری عملیات اجرا کنید.

۳. داده را به یک مقصد ارسال کنید.

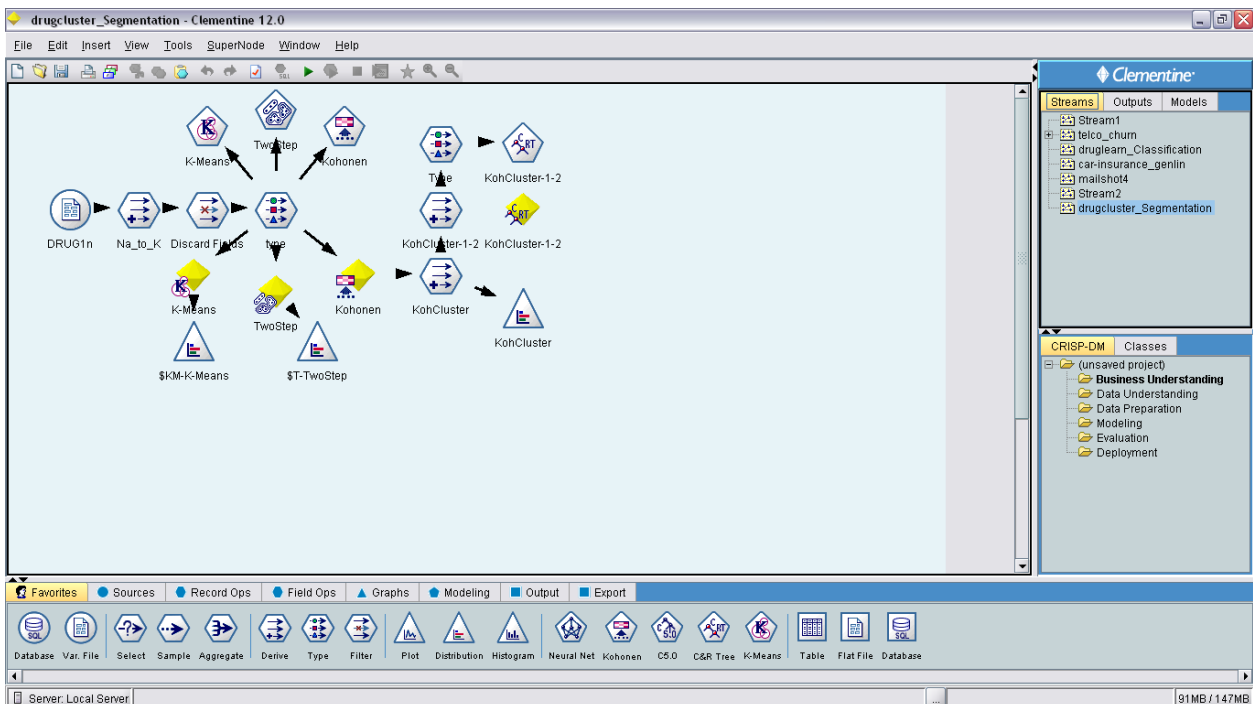
این ترتیب عملیات به عنوان «Data Stream» شناخته می‌شود.



## Stream Canvas

بزرگترین فضای موجود در پنجره نرم افزار که می‌توان Data Stream را در آن اجرا کرد. Stream ها را می‌توان با رسم دیاگرام‌های عملیات داده‌ها در Canvas ایجاد کرد.

## نمونه‌ای از Canvas:



## Node (گره)

هر عملیاتی به وسیله‌ی یک آیکن یا گره نشان داده می‌شوند. گره‌های متصل به هم نشان دهنده‌ی جریان داده‌ها هستند.

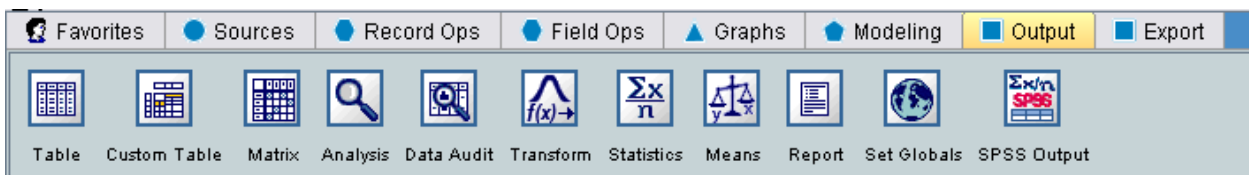
## Stream Manager

Stream ها در پنجره Stream Manager در گوشه بالا، سمت راست نرم افزار ذخیره می شوند.

### Node Palettes (جعبه گره ها)

جعبه ها شامل گره ها هستند که در قسمت پایین پنجره وجود دارند. شامل Favorites ، Source ، Record ops ، Graphs ، Field ops ، Modeling ، Output ، Export می باشد. وقتی روی نام یک جعبه کلیک می کنید، لیستی از گره های موجود در آن ظاهر می شوند.

مثلاً:



برای ایجاد گره در Canvas روی گره دو بار کلیک کنید و یا آن گره را به قسمت Canvas درگ کنید.

### انواع گره های موجود:

- Source: گره هایی که داده ها را وارد Clementine می کنند.
- Record ops: گره هایی که عملیاتی مانند ضبط، انتخاب، ادغام و ... را روی داده انجام می دهند.
- Field ops: گره هایی که عملیاتی مانند فیلتر کردن، ایجاد فیلد جدید و ... را روی داده ها انجام می دهند.
- Graphs: گره هایی که داده ها را قبل و بعد از مدل سازی به صورت گرافیکی نشان می دهند. این نمودارها شامل هیستوگرام و پراکندگی و ... می باشد.
- Modeling: گره هایی که از الگوریتم های موجود در مدل سازی استفاده می کنند. مثلاً شبکه عصبی، درخت تصمیم، الگوریتم خوشه بندی و ...
- Out put: گره هایی که صورت های مختلفی از خروجی داده مثل نمودارها، نتایج مدل ها و ... را به نحوی نشان می دهد که هم از طریق نرم افزار Clementine قابل خواندن باشد و هم نرم افزارهای دیگری مثل SPSS یا Excel.

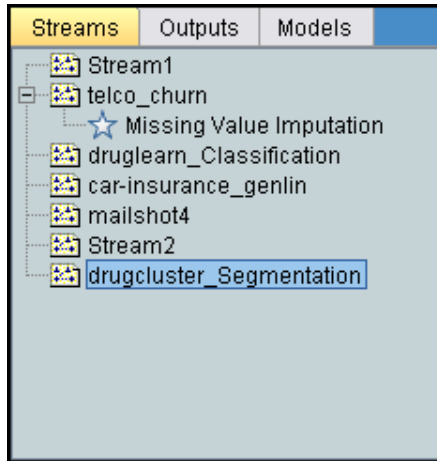
در قسمت Stream Manager، ۳ گزینه وجود دارد:

Streams -۱

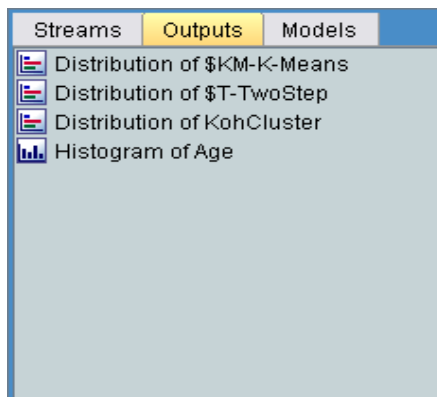
Outputs -۲

## Models -۳

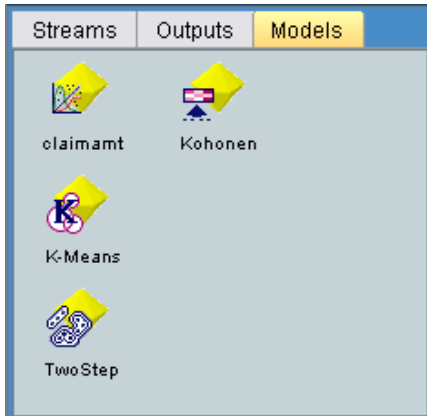
**Stream:** این قسمت، Stream های فعال را که با آنها می‌توان کار کرد، نشان می‌دهد. همچنین از این قسمت می‌توانید Stream های ایجاد شده را باز، تغییر نام، ذخیره و حذف کنید.







**Output:** در این بخش تمام جداول و نمودارها وجود دارند. شما می‌توانید این خروجی‌ها را باز، تغییر نام، ذخیره و حذف کنید.



**Model:** تمام مدل‌های ایجاد شده در این قسمت قابل نمایش است. این مدل‌ها از Model Tab می‌توانند باز و یا به Canvas اضافه شوند.



### ابزارهای Clementine :

	ایجاد Stream جدید		باز کردن Stream
	ذخیره سازی Stream		پرینت گرفتن از Stream حاضر
	باز کردن پوشه مخصوص مثال ها و Streamها		کات کردن
	کپی کردن		Paste کردن
	حرکت قبلی		دوباره انجام دادن
	ویرایش مشخصات Stream		اجرای Stream حاضر
	اجرای قسمت مشخصی از Stream		توقف اجرا
	اضافه کردن SuperNode		بزرگنمایی (مخصوص SuperNode)
	کوچک نمایی (مخصوص SuperNode)		

### اضافه کردن به پروژه:

زمانی که یک پروژه را ایجاد یا باز می کنید، می توانید، اشیاء مختلفی از جمله: Data Stream ، گره و .. را به پروژه اضافه کنید.

### اضافه کردن از طریق Managers:



از Managers در قسمت بالا و سمت راست پنجره می‌توان برای اضافه کردن به پروژه استفاده کرد. برای این کار:

- شیء را انتخاب کنید. (از جمله جدول و یا Stream و ...)
- راست کلیک کنید و "Add to Project" را انتخاب کنید. اگر آن شیء قبلاً ذخیره شده باشد، به صورت خودکار به پوشه‌ی مناسب اضافه می‌شود (در Class View) و یا به پوشه پیش فرض (در CRISP-DM View) اضافه می‌شود.

در غیر اینصورت می‌توانید شیء مورد نظر را به فضای کار پروژه درگ کنید.

### اضافه کردن گزاره Canvas:

برای این کار :

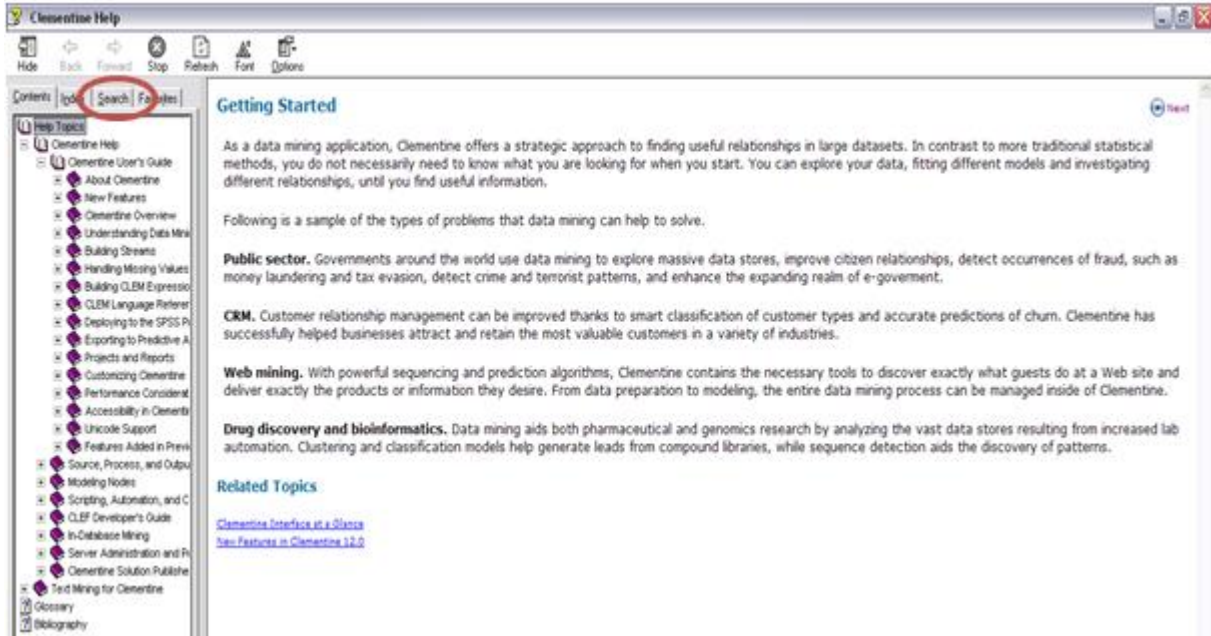
- گرهی را از Canvas انتخاب کنید.
- راست کلیک کرده و Save Node را انتخاب کنید در غیراینصورت از نوار منو مسیر `Edit > Node > Save` را طی کنید.
- در کادر مربوط "Add to Project" را انتخاب کنید.
- نامی برای گره در نظر بگیرید و آن را ذخیره کنید. (Save را انتخاب کنید)

### طریقه استفاده از Help نرم‌افزار:

همانطور که ملاحظه می‌کنید منوی Help شامل چند گزینه است که با توجه به کاری که می‌خواهید انجام دهید، می‌توانید از آن استفاده کنید.

### :Help Topics

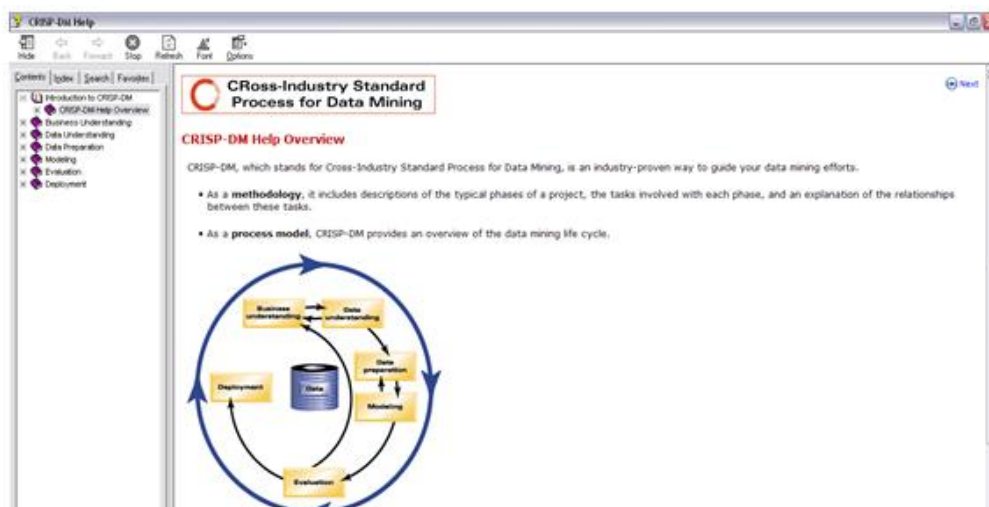
با استفاده از این گزینه می‌توانید موضوع مورد نظر خود را جست‌وجو کنید. این قسمت شامل توضیحاتی در مورد قسمت‌های مختلف نرم‌افزار، گره‌ها و ... است. همچنین Option‌های مختلف گره‌ها نیز در اینجا توضیح داده شده است.



از کادر سمت چپ می‌توانید موضوع مورد نظر خود را انتخاب کنید و یا آن را از قسمت Search (که در شکل نشان داده شده است) جست‌وجو کنید.

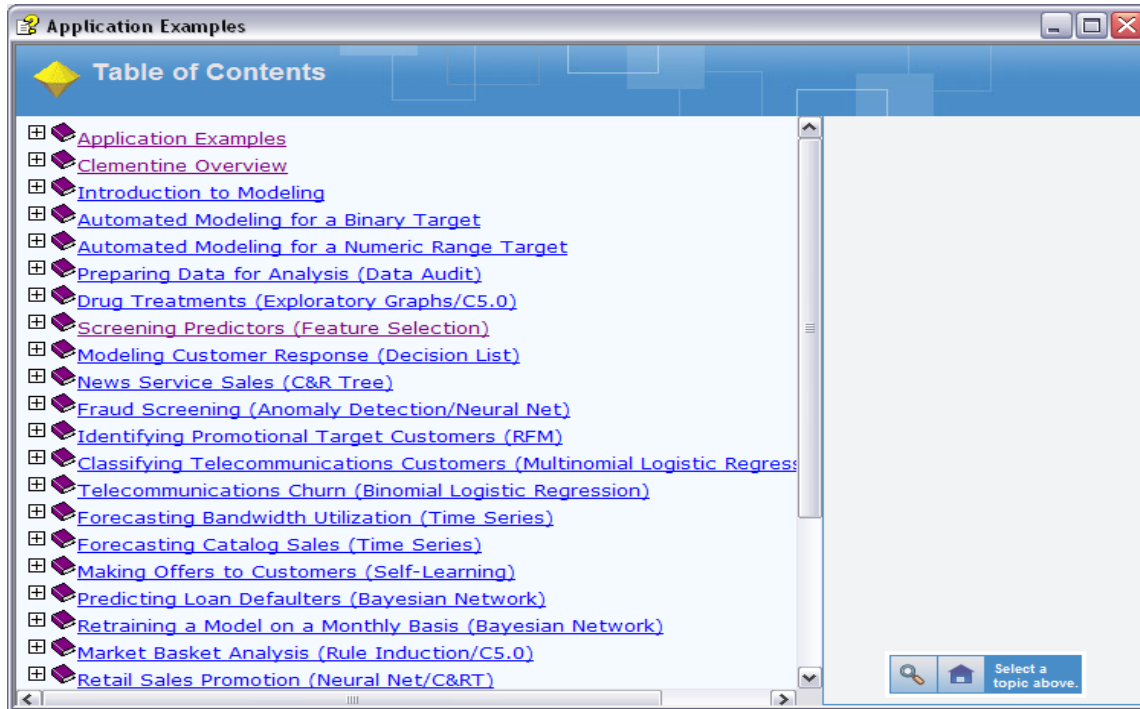
### :CRISP-DM Help

این گزینه اطلاعاتی در مورد مبحث داده‌کاوی، روش‌های مختلف آن، طرز برخورد با داده و سایر اطلاعات در این زمینه، در اختیاران قرار می‌دهد.



## :Application Examples

با استفاده از این گزینه می‌توانید به مثال‌های انجام شده دسترسی پیدا کنید و این مثال‌ها را قدم به قدم با نرم‌افزار انجام دهید. آدرس تمام مثال‌ها و Stream مورد استفاده نیز در این گزینه ذکر شده است.



## بخش دوم: کار با Stream ها

## مروری بر ایجاد Stream:

به فرآیند اداره کردن داده میان یک سری گره، Stream می‌گویند. این سری گره‌ها نشان دهنده‌ی عملیات بوده که بر روی داده انجام می‌شود. این در حالی است که ارتباط بین گره‌ها نشان دهنده‌ی مسیر جریان داده است.

به طور کلی، از Data Stream برای موارد زیر استفاده می‌شود:

- خواندن داده در Clementine
- کار با داده از طریق یک سری عملیات و سپس
- ارسال آن به مقصدی مانند یک فایل SPSS یا Clementine Solution Publisher.

برای مثال فرض کنید که قصد دارید:

- ۱- یک منبع ورودی داده را باز کنید.
- ۲- یک فیلد جدید اضافه کنید.
- ۳- رکوردهای موجود در فیلد جدید را انتخاب کنید و سپس
- ۴- نتایج را از طریق جدول نشان دهید.

در این مورد، Data Stream شما شامل ۴ گره خواهد بود.

Variable File Node، که از طریق آن داده‌هایتان را می‌خوانید.



Var. File

Derive Node، که به کمک آن یک فیلد جدید را به مجموعه داده اضافه می‌کنید.



Derive

Select Node، که با آن محدوده رکورد را برای شمول رکوردها از Data Stream، مشخص می‌شود.



Select

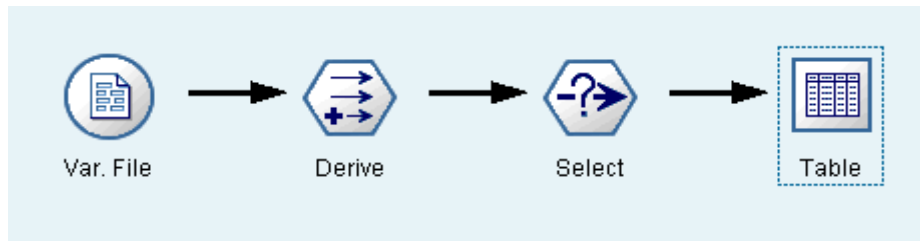
Table Node، که با استفاده از این گره، نتایج را بر روی صفحه نمایش، نشان می‌دهد.



Table

می‌توانید با کمک قدم‌های زیر یک Data Stream ایجاد کنید:

- اضافه کردن گره به Stream Canvas
- ایجاد ارتباط بین گره‌ها بمنظور ساخت Stream
- تعیین هر گونه گره یا گزینه‌های Stream
- اجرای Stream



### کار با گره‌ها :

سه راه برای اضافه کردن گره وجود دارد:

- روی گره دوبار کلیک کنید.
- گره را از جعبه گره به Stream، درگ کنید.
- روی گره در جعبه گره کلیک کنید و سپس روی فضای Stream Canvas کلیک کنید.

هنگامی که یک گره را به Stream Canvas اضافه کردید، روی گره دو بار کلیک کنید تا کادر مربوط به آن باز شود. گزینه‌های پیش روی شما به گرهی که انتخاب کردید، بستگی دارد. برای کسب اطلاعات در مورد کاربردهای کادر از Help نرم افزار کمک بگیرید.

برای حذف یک گره، روی آن کلیک و از صفحه کلید Delete را فشار دهید یا روی گره راست کلیک کنید و Delete را انتخاب کنید.

### ایجاد ارتباط بین گره‌ها:

گره‌هایی را که به Stream Canvas اضافه می‌شوند، تا زمانی که به یکدیگر متصل نشده‌اند یک Data Stream ایجاد نمی‌کنند. ارتباطات بین گره‌ها نشان دهنده‌ی مسیر جریان داده می‌باشد.

چند راه برای ایجاد این ارتباط بین گره ها وجود دارد:

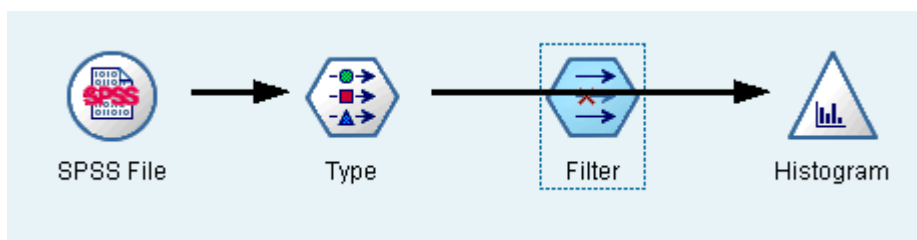
- دوبار کلیک (بعد از انتخاب یک گره، گره بعدی را انتخاب کنید)
- استفاده از کلید وسطی موس
- کلید سمت چپ موس + کلید Alt
- به صورت دستی (روی یک گره راست کلیک کرده و Connect را انتخاب کنید و سپس روی گره دیگر کلیک کنید).

در صورتیکه بخواهید هر یک از ارتباط های زیر را ایجاد کنید، یک پیغام خطا دریافت خواهید کرد:

- ارتباطی که به یک Source node ختم شود.
- ارتباطی که از یک Terminal node (گره پایانی) شروع شود.
- زمانی که به یک گره بیش از تعداد ظرفیتش ارتباط ایجاد کنید.
- اتصال دو گرهی که قبلاً با هم مرتبط شده اند.
- تشکیل شدن چرخه

### کنار گذاشتن گره در یک Stream:

زمانی که در Data Stream ، یک گره را کنار می گذارید، تمام ارتباطات ورودی و خروجیش با ارتباطی که بین گره ورودی و گره خروجی ایجاد می شود، جایگذاری می شود.



مشاهده می کنید که در این Stream، گره Filter کنار گذاشته شده است.

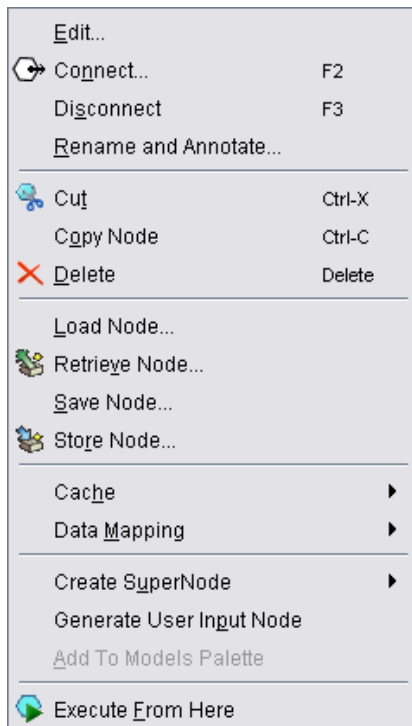
### حذف ارتباط:

- می توانید روی پیکان ارتباط، راست کلیک کرده و Delete Connection را انتخاب کنید.
- یک گره را انتخاب کنید و  $F_3$  را فشار دهید تا تمام ارتباطات حذف شوند.
- روی گره راست کلیک کنید و Disconnect را انتخاب کنید.
- مسیر زیر را دنبال کنید :

Edit &gt; Node &gt; Disconnect

**Option های یک گره:**

زمانی که یک گره راست کلیک می‌کنید می‌توانید به گزینه‌های مختلف آن دسترسی پیدا کنید.



**Edit:** برای باز کردن کادر مربوط به گره انتخابی

**Connect:** برای ایجاد ارتباط بین گره‌ها

**Disconnect:** برای حذف ارتباط بین گره‌ها

**Rename and annotation:** برای باز کردن کادر Annotation و

تغییر نام و ...

**Cut:** برای حذف گره در حالی که می‌توان بعداً Paste کرد.

**Delete:** حذف دائمی گره

**Copy:** برای کپی کردن گره

**Load Node:** برای باز کردن گره‌ی که قبلاً ذخیره شده و بارگذاری

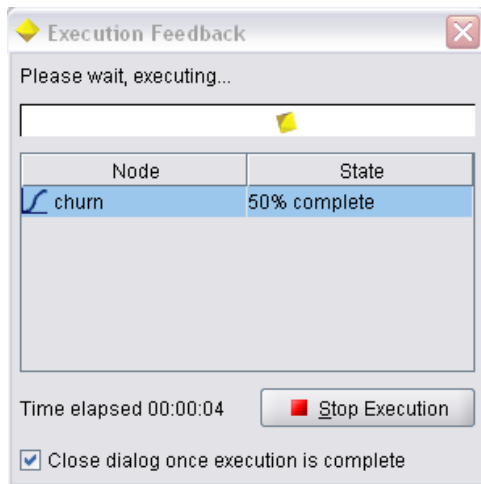
کردن Option های آن در گره‌ی که هم اکنون انتخاب شده است.

**Save Node:** برای ذخیره کردن گره.

**اجرای Stream ها:**

در Clementine چند راه برای اجرای Stream وجود دارد:

- می‌توانید از منوی «Tools» گزینه‌ی «Execute» را انتخاب کنید.
- از کلید مربوط به اجرا در نوار ابزار می‌توانید کمک بگیرید.
- روی گره نهایی راست کلیک کرده و Execute را انتخاب کنید.
- روی یک گره راست کلیک کرده و Execute From Here را انتخاب کنید.



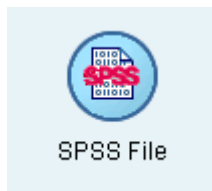
برای متوقف کردن یک Stream در حال اجرا می‌توانید روی دکمه قرمز در نوار ابزار کلیک کنید و یا گزینه‌ی Stop Execute را از منوی Tools انتخاب کنید.

اگر زمان اجرا از ۳ ثانیه بیشتر طول کشید، کادر Execute Feedback ظاهر می‌شود و پیشرفت اجرا را نمایش می‌دهد.

## کار با انواع داده

### اضافه کردن گره Source:

گره SPSS را انتخاب کرده و فایل telco.sav را برای آن در نظر بگیرید.



### اضافه کردن گره Type:

گره Type را از بخش Field ops در جعبه گره انتخاب کرده و به گره SPSS متصل کنید.









مشخصات مربوط به یک فیلد هم می‌تواند در گره Source و هم در گره Type تعیین شوند.

## Data Type (نوع داده)



- طریقه‌ی استفاده از فیلدهای داده‌ها را در Clementine توضیح می‌دهد.
- مشخصه‌ی داده‌های یک فیلد را تعیین می‌کند.
- اگر تمام جزئیات یک فیلد معلوم بود، آن فیلد Fully instantiated (به طور کامل معرفی شده) نام دارد.
- نوع یک فیلد با نحوه ذخیره آن متفاوت است. داده‌ها می‌توانند به صورت رشته‌ای (String)، عدد صحیح (integers)، تاریخ، زمان ذخیره شوند. مثلاً ممکن است بخواهید، Type یک داده‌ی عددی (integers) را به صورت Flag تعریف کنید.

### حالات مختلف Data Type:

- Range : برای استفاده از داده‌های عددی به صورت یک دامنه مثل 0-100 که این اعداد می‌توانند: صحیح، حقیقی و ... باشند.
- Discrete : برای مقادیر رشته‌ای (String) و زمانی که با اعداد گسسته سر و کار داریم به کار می‌رود.
- Flag : زمانی که تنها دو مقدار ممکن برای فیلد مورد نظر شما موجود است، مثل «M , F»
- Set : زمانی که داده چند مقدار را می‌تواند به خود بگیرد: مثل Low / Normal / High و ...
- Ordered Set : کاربرد این نوع داده‌ها همانند Set می‌باشد با این تفاوت که مقادیر اختصاص داده شده به یک فیلد باید دارای یک ترتیب خاص باشند. (مثلاً صعودی یا به صورت الفبا)
- Typeless : اگر داده‌ای با هیچکدام از Type‌های فوق مطابقت نداشته، آنگاه در این بخش جای می‌گیرد. اگر Type داده‌ای را به صورت Type less انتخاب کنید به طور خودکار نقش (role) آن فیلد به صورت None تغییر خواهد کرد.


### :Values


- <Read>: زمانی که گره را اجرا کنید، داده‌ها تنها خوانده می‌شوند.
- <Pass>: هیچ داده‌ای خوانده نمی‌شود.
- <Current>: مقدار فعلی داده را نگه می‌دارد.
- <Specify...>: کادر دیگری باز می‌شود تا در آن مشخصات فیلد را تعیین نمایید.


زمانی که دکمه Read Values را انتخاب کنید، مقادیر فیلدهای موجود، نمایش داده می‌شوند.


### :Direction

این ستون نقش فیلد را در عملیات داده‌کاوی معین می‌کند. یک فیلد می‌تواند به عنوان ورودی، هدف یا خروجی، هم ورودی و هم خروجی شناخته شود و یا اصلاً هیچ نقشی در پروژه‌ی داده‌کاوی نداشته باشد.


 In : بعنوان متغیر ورودی و یا پیش‌بینی کننده، فیلد مورد نظر را معرفی می‌کند.

 Out : فیلد را بعنوان متغیر هدف تعیین می‌نماید.

 Both : با انتخاب این گزینه فیلد هم ورودی عملیات است و هم هدف.

 None : اگر این حالت انتخاب شود دیگر آن فیلد نقشی در کار نخواهد داشت. (اگر Type یک فیلد بصورت

Typeless انتخاب شود، به طور پیش‌فرض Direction آن فیلد به None تغییر پیدا می‌کند.)

 Partition : این گزینه برای استفاده از یک فیلد برای دسته‌بندی داده‌ها به کار می‌رود. فیلد مورد نظر باید دو یا سه مقدار داشته باشد.