Fire Up GraphLab Create [Created by Ranjeet Singh Mahla]

In [1]:
```
import graphlab
```

Laod a Tabular Data Set; comman seperated file csv

In [3]:
```
sf =graphlab.SFrame('people-example.csv')
```

```
[INFO] graphlab.cython.cy_server: GraphLab Create v2.1 started. Loggin
g: /tmp/graphlab_server_1517906182.log

This non-commercial license of GraphLab Create for academic use is ass
igned to ranjeet@iiserb.ac.in and will expire on January 09, 2019.

Finished parsing file /Users/ranjeet/Desktop/people-example.csv

Parsing completed. Parsed 7 lines in 0.025882 secs.

------------------------------------------------------
Inferred types from first 100 line(s) of file as
column_type_hints=[str,str,str,int]
If parsing fails due to incorrect types, you can correct
the inferred type list above and pass it to read_csv in
the column_type_hints argument
------------------------------------------------------

Finished parsing file /Users/ranjeet/Desktop/people-example.csv

Parsing completed. Parsed 7 lines in 0.012891 secs.
```

SFrame Basics

In [4]:
```
sf
```

Out[4]:

First Name	Last Name	Country	age
Bob	Smith	United States	24
Alice	Williams	Canada	23
Malcolm	Jone	England	22
Felix	Brown	USA	23
Alex	Cooper	Poland	23
Tod	Campbell	United States	22
Derek	Ward	Switzerland	25

[7 rows x 4 columns]

In [5]:
```
sf.head() # frist few line of data set
```

Out[5]:

First Name	Last Name	Country	age
Bob	Smith	United States	24
Alice	Williams	Canada	23
Malcolm	Jone	England	22
Felix	Brown	USA	23
Alex	Cooper	Poland	23
Tod	Campbell	United States	22
Derek	Ward	Switzerland	25

[7 rows x 4 columns]

In [6]:
```
sf.tail() # last few line of data set
```

Out[6]:

First Name	Last Name	Country	age
Bob	Smith	United States	24
Alice	Williams	Canada	23
Malcolm	Jone	England	22
Felix	Brown	USA	23
Alex	Cooper	Poland	23
Tod	Campbell	United States	22
Derek	Ward	Switzerland	25

[7 rows x 4 columns]

Graphlab Canvas

In [7]:
```
sf.show()
```

```
Canvas is accessible via web browser at the URL: http://localhost:5190
0/index.html (http://localhost:51900/index.html)
Opening Canvas in default web browser.
```

In [8]:
```
graphlab.canvas.set_target('ipynb')
```

In [9]:
```
sf['age'].show(view ='Categorical')
```

Most frequent items from <*SArray*>

Value	Count	Percent	
23	3	42.857%	
22	2	28.571%	
24	1	14.286%	
25	1	14.286%	

Inspect Column of Dataset

In [10]:
```
sf['Country']
```

Out[10]:
```
dtype: str
Rows: 7
['United States', 'Canada', 'England', 'USA', 'Poland', 'United States
', 'Switzerland']
```

In [11]:
```
sf['age']
```

Out[11]:
```
dtype: int
Rows: 7
[24, 23, 22, 23, 23, 22, 25]
```

In [12]:
```
sf['age'].mean()
```

Out[12]:
```
23.142857142857146
```

In [14]:
```
sf['age'].min()
```

Out[14]:
```
22
```

In [15]:
```
sf['age'].max()
```

Out[15]:
```
25
```

Creation of New Column in Data Frame

In [16]:
```
sf
```

Out[16]:

First Name	Last Name	Country	age
Bob	Smith	United States	24
Alice	Williams	Canada	23
Malcolm	Jone	England	22
Felix	Brown	USA	23
Alex	Cooper	Poland	23
Tod	Campbell	United States	22
Derek	Ward	Switzerland	25

[7 rows x 4 columns]

In [21]:
```
sf['Full Name'] = sf['First Name']+ ' ' +sf['Last Name']
```

In [20]:
```
sf
```

Out[20]:

First Name	Last Name	Country	age	Full Name
Bob	Smith	United States	24	Bob Smith
Alice	Williams	Canada	23	Alice Williams
Malcolm	Jone	England	22	Malcolm Jone
Felix	Brown	USA	23	Felix Brown
Alex	Cooper	Poland	23	Alex Cooper
Tod	Campbell	United States	22	Tod Campbell
Derek	Ward	Switzerland	25	Derek Ward

[7 rows x 5 columns]

Few Statics with data frame

In [22]:
```
sf['age']+2
```

Out[22]:
```
dtype: int
Rows: 7
[26, 25, 24, 25, 25, 24, 27]
```

In [23]:
```
sf
```

Out[23]:

First Name	Last Name	Country	age	Full Name
Bob	Smith	United States	24	Bob Smith
Alice	Williams	Canada	23	Alice Williams
Malcolm	Jone	England	22	Malcolm Jone
Felix	Brown	USA	23	Felix Brown
Alex	Cooper	Poland	23	Alex Cooper
Tod	Campbell	United States	22	Tod Campbell
Derek	Ward	Switzerland	25	Derek Ward

[7 rows x 5 columns]

In [24]:
```
sf['age']*sf['age']
```

Out[24]:
```
dtype: int
Rows: 7
[576, 529, 484, 529, 529, 484, 625]
```

In [25]:
```
sf['age']-sf['age']
```

Out[25]:
```
dtype: int
Rows: 7
[0, 0, 0, 0, 0, 0, 0]
```

Data transformation using apply() function

In [26]:
```
sf['Country'].show()
```

Most frequent items from <*SArray*>

Value	Count	Percent
United States	2	28.571%
Canada	1	14.286%
England	1	14.286%
USA	1	14.286%
Poland	1	14.286%
Switzerland	1	14.286%

In [28]:
```
def transform_country(country):
    if country == 'USA':
        return 'United States'
    else:
        return country
```

In [29]:
```
sf
```

Out[29]:

First Name	Last Name	Country	age	Full Name
Bob	Smith	United States	24	Bob Smith
Alice	Williams	Canada	23	Alice Williams
Malcolm	Jone	England	22	Malcolm Jone
Felix	Brown	USA	23	Felix Brown
Alex	Cooper	Poland	23	Alex Cooper
Tod	Campbell	United States	22	Tod Campbell
Derek	Ward	Switzerland	25	Derek Ward

[7 rows x 5 columns]

In [30]:
```
transform_country('Brazil')
```

Out[30]:
```
'Brazil'
```

In [31]:
```
transform_country('USA')
```

Out[31]:
```
'United States'
```

In [32]:
```
def transform_country(country):
    if country == 'Brazil':
        return 'Brasil'
    else:
        return country
```

In [33]:
```
sf
```

Out[33]:

First Name	Last Name	Country	age	Full Name
Bob	Smith	United States	24	Bob Smith
Alice	Williams	Canada	23	Alice Williams
Malcolm	Jone	England	22	Malcolm Jone
Felix	Brown	USA	23	Felix Brown
Alex	Cooper	Poland	23	Alex Cooper
Tod	Campbell	United States	22	Tod Campbell
Derek	Ward	Switzerland	25	Derek Ward

[7 rows x 5 columns]

In [36]:
```
sf['Country'] = sf['Country'].apply(transform_country) # The command chang
```

In [ ]:
```

```