If you're already familiar with Colab, check out this video to learn about interactive tables, the executed code history view and the command palette.

Colab, or 'Colaboratory', allows you to write and execute Python in your browser, with

- Zero configuration required
- Access to GPUs free of charge
- Easy sharing

Whether you're a student, a data scientist or an AI researcher, Colab can make your work easier. Watch Introduction to Colab to find out more, or just get started below!

The document that you are reading is not a static web page, but an interactive environment called a Colab notebook that lets you write and execute code.

For example, here is a code cell with a short Python script that computes a value, stores it in a variable and prints the result:

```
seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day
```

```
86400
```

To execute the code in the above cell, select it with a click and then either press the play button to the left of the code, or use the keyboard shortcut 'Command/Ctrl+Enter'. To edit the code, just click the cell and start editing.

Variables that you define in one cell can later be used in other cells:

```
seconds_in_a_week = 7 * seconds_in_a_day
seconds_in_a_week
```

```
604800
```

Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, LaTeX and more. When you create your own Colab notebooks, they are stored in your Google Drive account. You can easily share your Colab notebooks with co-workers or friends, allowing them to comment on your notebooks or even edit them. To find out more, see Overview of Colab. To create a new Colab notebook you can use the File menu above, or use the following link: Create a new Colab notebook.

Colab notebooks are Jupyter notebooks that are hosted by Colab. To find out more about the Jupyter project, see jupyter.org.

With Colab you can harness the full power of popular Python libraries to analyse and visualise data. The code cell below uses numpy to generate some random data, and uses matplotlib to visualise it. To edit the code, just click the cell and start editing.

```
import numpy as np
import IPython.display as display
from matplotlib import pyplot as plt
import io
import base64

ys = 200 + np.random.randn(100)
x = [x for x in range(len(ys))]

fig = plt.figure(figsize=(4, 3), facecolor='w')
plt.plot(x, ys, '-')
plt.fill_between(x, ys, 195, where=(ys > 195), facecolor='g',
alpha=0.6)
plt.title("Sample Visualization", fontsize=10)

data = io.BytesIO()
plt.savefig(data)
image = F"
{base64.b64encode(data.getvalue()).decode()}"
alt = "Sample Visualization"
display.display(display.Markdown(F"""![{alt}]({image})"""))
plt.close(fig)
```

```
<IPython.core.display.Markdown object>
```

You can import your own data into Colab notebooks from your Google Drive account, including from spreadsheets, as well as from GitHub and many other sources. To find out more about importing data, and how Colab can be used for data science, see the links below under Working with data.

With Colab you can import an image dataset, train an image classifier on it, and evaluate the model, all in just a few lines of code. Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including GPUs and TPUs, regardless of the power of your machine. All you need is a browser.

Colab is used extensively in the machine learning community with applications including:

- Getting started with TensorFlow
- Developing and training neural networks
- Experimenting with TPUs
- Disseminating AI research
- Creating tutorials

To see sample Colab notebooks that demonstrate machine learning applications, see the machine learning examples below.

- Overview of Colaboratory
- Guide to markdown
- Importing libraries and installing dependencies
- Saving and loading notebooks in GitHub

- Interactive forms
- Interactive widgets
- Loading data: Drive, Sheets and Google Cloud Storage
- Charts: visualising data
- Getting started with BigQuery

Machine learning crash course

These are a few of the notebooks from Google's online machine learning course. See the full course website for more.

- Intro to Pandas DataFrame
- Linear regression with tf.keras using synthetic data
- TensorFlow with GPUs
- TensorFlow with TPUs
- NeMo voice swap: Use Nvidia NeMo conversational AI toolkit to swap a voice in an audio fragment with a computer-generated one.
- Retraining an Image Classifier: Build a Keras model on top of a pre-trained image classifier to distinguish flowers.
- Text Classification: Classify IMDB film reviews as either positive or negative.
- Style Transfer: Use deep learning to transfer style between images.
- Multilingual Universal Sentence Encoder Q&A: Use a machine-learning model to answer questions from the SQuAD dataset.
- Video Interpolation: Predict what happened in a video between the first and the last frame.

```
import pandas as pd
df=pd.DataFrame({"a":[4,5,6],
                 "b":[7,8,9],
                 "c":[1,2,3]},
                index=[1,2,3])
print(df)
```

```
   a  b  c
1  4  7  1
2  5  8  2
3  6  9  3
```

```
import pandas as pd
data=[["alex",10],["bob",12]]
df=pd.DataFrame(data,columns=["Name","Age"])
print(df)
```

```
   Name  Age
0  alex   10
1   bob   12
```

```
import pandas as pd
mydataset={"cars":["benz","volvo"],"passing":[2,3]}
myvar=pd.DataFrame(mydataset)
print(myvar)
```

```
    cars  passing
0   benz        2
1  volvo        3
```

```
import pandas as pd
data=[{"a":1,"b":2,"c":3},{"a":5,"b":6,"c":8}]
df=pd.DataFrame(data,index=["one","two"])
df
```


```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"a\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 5,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          5,\n          1\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"b\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 2,\n        \"min\": 2,\n
\"max\": 6,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          6,\n          2\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"c\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 3,\n        \"min\": 3,\n
\"max\": 8,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          8,\n          3\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```


```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],"Height":
[5.1,5.2,6.4,4.8],"Qualification":["MSC","MA","MBA","MCA"]}
df=pd.DataFrame(data,index=["i","ii","iii","iv"])
address=["Delhi","Chennai","Kerala","Patna"]
df["Address"]=address
print(df)
```

```
       Name  Height Qualification  Address
i       Jai     5.1           MSC    Delhi
ii   Princi     5.2            MA  Chennai
iii  Gaurav     6.4           MBA   Kerala
iv     Anuj     4.8           MCA    Patna
```

```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],"Height":
[5.1,5.2,6.4,4.8],"Qualification":["MSC","MA","MBA","MCA"],
      "address":["Delhi","Chennai","Kerala","Patna"]}
df=pd.DataFrame(data,index=["i","ii","iii","iv"])
del df["address"]
print(df)
```

```
       Name  Height Qualification
i       Jai     5.1           MSC
ii   Princi     5.2            MA
iii  Gaurav     6.4           MBA
iv     Anuj     4.8           MCA
```

```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],"Height":
[5.1,5.2,6.4,4.8],"Qualification":["MSC","MA","MBA","MCA"],
      "address":["Delhi","Chennai","Kerala","Patna"]}
df=pd.DataFrame(data,index=["i","ii","iii","iv"])
df.drop(["Height"],axis=1,inplace=True)
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 4,\n
\"samples\": [\n          \"Princi\",\n          \"Anuj\",\n
\"Jai\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Qualification\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          \"MA\",\n          \"MCA\",\n          \"MSC\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"address\",\n      \"properties\": {\n        \"dtype\": \"string\",\
n        \"num_unique_values\": 4,\n        \"samples\": [\n
\"Chennai\",\n          \"Patna\",\n          \"Delhi\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df.rename(columns={"address":"place"},inplace=True)
print(df)
```

```
    Name  Height Qualification    place
0    Jai     5.1           MSC    Delhi
1  Princi     5.2            MA  Chennai
```

```
import pandas as pd
df=pd.DataFrame([[1,2],[3,4]],columns=["a","b"])
df1=pd.DataFrame([[5,6],[7,8]],columns=["a","b"])
df=pd.concat([df,df1])
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"a\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 1,\n
\"max\": 7,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          3,\n          7,\n          1\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"b\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 2,\n
\"max\": 8,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          4,\n          8,\n          2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"df"}
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df.drop(0,axis=0,inplace=True)
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 1,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 1,\n
\"samples\": [\n          \"Princi\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\": null,\n
\"min\": 5.2,\n        \"max\": 5.2,\n        \"num_unique_values\":
1,\n        \"samples\": [\n          5.2\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Qualification\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n          \"MA\"\n
],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n
}\n    },\n    {\n      \"column\": \"address\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 1,\n        \"samples\": [\n
\"Chennai\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```

```
#indexing
import pandas as pd
data={"name":["Alice","Bob"],"age":[24,21],"gender":
```

```
["F","M"],"height":[2.5,2.1]}
df=pd.DataFrame(data)
df=df['height']
df
```

```
0    2.5
1    2.1
Name: height, dtype: float64
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df=df[["Name","Height"]]
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Princi\",\n          \"Jai\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.07071067811865513,\n        \"min\": 5.1,\n        \"max\": 5.2,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          5.2,\n
5.1\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```

```
#filter
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df.filter(items=['Name','Height'])
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Princi\",\n          \"Jai\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.07071067811865513,\n        \"min\": 5.1,\n        \"max\": 5.2,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          5.2,\n
5.1\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
```

```
df=pd.DataFrame(data)
df.filter(like="e")
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Princi\",\n          \"Jai\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.07071067811865513,\n        \"min\": 5.1,\n        \"max\": 5.2,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          5.2,\n
5.1\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Address\",\n      \"properties\": {\n        \"dtype\": \"string\",\
n        \"num_unique_values\": 2,\n        \"samples\": [\n
\"Chennai\",\n          \"Delhi\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df.filter(regex="e|a",axis=1)
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"Name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Princi\",\n          \"Jai\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.07071067811865513,\n        \"min\": 5.1,\n        \"max\": 5.2,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          5.2,\n
5.1\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"Qualification\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"MA\",\n          \"MSC\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"Address\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Chennai\",\n          \"Delhi\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
#deleting duplicates
import pandas as pd
data={"name":["Alice","Bob","Alice","Bob"],
```

```
      "age":[24,21,22,25],
      "gender":["F","M","M","F"],
      "height":[2.5,2.1,2.3,2.6]}
df=pd.DataFrame(data)
df.drop_duplicates()
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Bob\",\n          \"Alice\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 1,\n        \"min\": 21,\n
\"max\": 25,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          21,\n          25\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"gender\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"M\",\n          \"F\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.2217355782608345,\n        \"min\": 2.1,\n        \"max\": 2.6,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          2.1,\n
2.6\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"name":["Alice","Bob","Alice","Bob"],
      "age":[24,21,22,25],
      "gender":["F","M","M","F"],
      "height":[2.5,2.1,2.3,2.6]}
df=pd.DataFrame(data)
df.drop_duplicates(subset=["name"])
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Bob\",\n          \"Alice\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 21,\n
\"max\": 24,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          21,\n          24\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"gender\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"M\",\n          \"F\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
```

{\n   \"dtype\": \"number\",\n   \"std\":
0.28284271247461895,\n   \"min\": 2.1,\n   \"max\": 2.5,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   2.1,\n
2.5\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
import pandas as pd
data={"name":["Alice","Bob","Alice","Bob"],
      "age":[22,21,22,25],
      "gender":["F","M","M","F"],
      "height":[2.5,2.1,2.3,2.6]}
df=pd.DataFrame(data)
df.drop_duplicates(subset=["name","age"])
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 3,\n \"fields\": [\n
{\n   \"column\": \"name\",\n   \"properties\": {\n
\"dtype\": \"string\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"Bob\",\n   \"Alice\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\n
\"dtype\": \"number\",\n   \"std\": 2,\n   \"min\": 21,\n
\"max\": 25,\n   \"num_unique_values\": 3,\n   \"samples\":
[\n   22,\n   21\n   ],\n   \"semantic_type\":
\"\",\n   \"description\": \"\"\n   }\n   },\n   {\n
\"column\": \"gender\",\n   \"properties\": {\n   \"dtype\":
\"string\",\n   \"num_unique_values\": 2,\n   \"samples\":
[\n   \"M\",\n   \"F\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"height\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
0.264575131106459,\n   \"min\": 2.1,\n   \"max\": 2.6,\n
\"num_unique_values\": 3,\n   \"samples\": [\n   2.5,\n
2.1\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
import pandas as pd
data={"name":["Alice","Bob","Alice","Bob"],
      "age":[22,21,22,25],
      "gender":["F","M","M","F"],
      "height":[2.5,2.1,2.3,2.6]}
df=pd.DataFrame(data)
df.drop_duplicates(subset=["name","age"],keep="last")
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 3,\n \"fields\": [\n
{\n   \"column\": \"name\",\n   \"properties\": {\n
\"dtype\": \"string\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"Alice\",\n   \"Bob\"\n   ],\n

```
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 21,\n
\"max\": 25,\n        \"num_unique_values\": 3,\n        \"samples\":
[\n          21,\n          22\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"gender\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          \"F\",\n          \"M\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.2516611478423583,\n        \"min\": 2.1,\n        \"max\": 2.6,\n
\"num_unique_values\": 3,\n        \"samples\": [\n          2.1,\n
2.3\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df_sample=df.sample(n=1)
print(df_sample)
```

```
  Name  Height Qualification Address
0  Jai     5.1           MSC   Delhi
```

```
import pandas as pd
data={"Name":["Jai","Princi"],"Height":[5.1,5.2],"Qualification":
["MSC","MA"],"Address":["Delhi","Chennai"]}
df=pd.DataFrame(data)
df_sample=df.sample(n=1)
print(df_sample)
```

```
     Name  Height Qualification  Address
1  Princi     5.2            MA  Chennai
```

```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],
      "Height":[5.1,5.2,6.4,4.8],
      "Qualification":["MSC","MA","MBA","MCA"],
      "address":["Delhi","Chennai","Kerala","Patna"]}
df=pd.DataFrame(data)
df_sample=df.sample(frac=0.5)
print(df_sample)
```

```
   Name  Height Qualification address
3  Anuj     4.8           MCA   Patna
0   Jai     5.1           MSC   Delhi
```

```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],
      "Height":[5.1,5.2,6.4,4.8],
      "Qualification":["MSC","MA","MBA","MCA"],
      "address":["Delhi","Chennai","Kerala","Patna"]}
df=pd.DataFrame(data)
df_sample=df.sample(n=2,axis=1)
print(df_sample)
```

```
     Name Qualification
0     Jai           MSC
1  Princi            MA
2  Gaurav           MBA
3    Anuj           MCA
```

```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],
      "age":[23,25,22,26],
      "salary":[20000,30000,18000,25000]
      }
df=pd.DataFrame(data)
high_salary=df.nlargest(2,columns="salary")
high_salary
```


```
{"summary":"{\n  \"name\": \"high_salary\",\n  \"rows\": 2,\n
\"fields\": [\n    {\n      \"column\": \"Name\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Anuj\",\n          \"Princi\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 0,\n        \"min\": 25,\n
\"max\": 26,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          26,\n          25\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"salary\",\n      \"properties\": {\n        \"dtype\":
\"number\",\n        \"std\": 3535,\n        \"min\": 25000,\n
\"max\": 30000,\n        \"num_unique_values\": 2,\n
\"samples\": [\n          25000,\n          30000\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe","variable_name":"high_salary"}
```


```
import pandas as pd
data={"Name":["Jai","Princi","Gaurav","Anuj"],
      "age":[23,25,22,26],
      "salary":[20000,30000,18000,25000]
      }
df=pd.DataFrame(data)
high_salary=df.nsmallest(2,columns="salary")
high_salary
```

{"summary":"{\n \"name\": \"high_salary\",\n \"rows\": 2,\n
\"fields\": [\n   {\n   \"column\": \"Name\",\n
\"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 2,\n   \"samples\": [\n   \"Jai\",\
n   \"Gaurav\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 0,\n   \"min\": 22,\n   \"max\": 23,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   23,\n
22\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"salary\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 1414,\n   \"min\": 18000,\n   \"max\": 20000,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   20000,\n
18000\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\
n}","type":"dataframe","variable_name":"high_salary"}

```
import pandas as pd
data={"name":["Alice","Bob","Dave","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df=df.query("age>=30")
df
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 2,\n \"fields\": [\n
{\n   \"column\": \"name\",\n   \"properties\": {\n
\"dtype\": \"string\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"Charlie\",\n   \"Alice\"\
n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 2,\n   \"min\": 31,\n   \"max\": 34,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   34,\n
31\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"gender\",\n   \"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 1,\n   \"samples\": [\n   \"F\"\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"height\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
0.6363961030678927,\n   \"min\": 1.6,\n   \"max\": 2.5,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   1.6\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   }\n]\n}","type":"dataframe","variable_name":"df"}

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
```

```
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df=df.query('name.str.contains("a") and height<=2')
df
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 2,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"Charlie\",\n          \"Dave\"\
n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 9,\n        \"min\": 21,\n        \"max\": 34,\n
\"num_unique_values\": 2,\n        \"samples\": [\n          34,\n
21\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"gender\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"F\",\n
\"M\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"height\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.28284271247461895,\n        \"min\": 1.6,\n        \"max\":
2.0,\n        \"num_unique_values\": 2,\n        \"samples\": [\n
1.6,\n          2.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.loc[:,"age"]
```

```
0    31
1    21
2    22
3    34
Name: age, dtype: int64
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.iloc[:,1]
```

```
0    31
1    21
2    22
3    34
Name: age, dtype: int64
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.loc[:,["name","age"]]
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 4,\n
\"samples\": [\n          \"Dave\",\n          \"Charlie\",\n
\"Alice\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 6,\n        \"min\": 21,\n        \"max\": 34,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          21,\n
34,\n          31\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df_filtered=df[df["age"]>30]
df_filtered
```

```
{"summary":"{\n  \"name\": \"df_filtered\",\n  \"rows\": 2,\n
\"fields\": [\n    {\n      \"column\": \"name\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n
\"Charlie\",\n          \"Alice\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 2,\n        \"min\": 31,\n
\"max\": 34,\n        \"num_unique_values\": 2,\n        \"samples\":
[\n          34,\n          31\n        ],\n        \"semantic_type\":
\"\",\n        \"description\": \"\"\n      }\n    },\n    {\n
\"column\": \"gender\",\n      \"properties\": {\n        \"dtype\":
\"string\",\n        \"num_unique_values\": 1,\n        \"samples\":
[\n          \"F\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
```

\"height\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": 0.6363961030678927,\n   \"min\": 1.6,\n   \"max\":
2.5,\n   \"num_unique_values\": 2,\n   \"samples\": [\n
1.6\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\
n}","type":"dataframe","variable_name":"df_filtered"}

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df_filtered=df[(df["gender"]=="M") & (df["height"]>=2.3)]
df_filtered
```

{"summary":"{\n \"name\": \"df_filtered\",\n \"rows\": 1,\n
\"fields\": [\n   {\n   \"column\": \"name\",\n
\"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 1,\n   \"samples\": [\n   \"Bob\"\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": null,\n
\"min\": 22,\n   \"max\": 22,\n   \"num_unique_values\": 1,\
n   \"samples\": [\n   22\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"gender\",\n   \"properties\":
{\n   \"dtype\": \"string\",\n   \"num_unique_values\": 1,\n
\"samples\": [\n   \"M\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"height\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\": null,\n
\"min\": 2.3,\n   \"max\": 2.3,\n   \"num_unique_values\":
1,\n   \"samples\": [\n   2.3\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df_filtered"}

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df_filtered=df[df["name"].str.startswith(("A"))]
df_filtered
```

{"summary":"{\n \"name\": \"df_filtered\",\n \"rows\": 1,\n
\"fields\": [\n   {\n   \"column\": \"name\",\n
\"properties\": {\n   \"dtype\": \"string\",\n
\"num_unique_values\": 1,\n   \"samples\": [\n

\"Alice\"\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   },\n   {\n   \"column\":
\"age\",\n   \"properties\": {\n   \"dtype\": \"number\",\n
\"std\": null,\n   \"min\": 31,\n   \"max\": 31,\n
\"num_unique_values\": 1,\n   \"samples\": [\n   31\n
],\n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"gender\",\n   \"properties\":
{\n   \"dtype\": \"string\",\n   \"num_unique_values\": 1,\n
\"samples\": [\n   \"F\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"height\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\": null,\n
\"min\": 2.5,\n   \"max\": 2.5,\n   \"num_unique_values\":
1,\n   \"samples\": [\n   2.5\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   }\n]\n}","type":"dataframe","variable_name":"df_filtered"}

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6],
      }
df=pd.DataFrame(data)
df.tail(2)
```

{"summary":"{\n \"name\": \"df\",\n \"rows\": 2,\n \"fields\": [\n
{\n   \"column\": \"name\",\n   \"properties\": {\n
\"dtype\": \"string\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"Charlie\",\n   \"Bob\"\n   ],\
n   \"semantic_type\": \"\",\n   \"description\": \"\"\n
}\n   },\n   {\n   \"column\": \"age\",\n   \"properties\": {\
n   \"dtype\": \"number\",\n   \"std\": 8,\n   \"min\":
22,\n   \"max\": 34,\n   \"num_unique_values\": 2,\n
\"samples\": [\n   34,\n   22\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"gender\",\n   \"properties\":
{\n   \"dtype\": \"string\",\n   \"num_unique_values\": 2,\n
\"samples\": [\n   \"F\",\n   \"M\"\n   ],\n
\"semantic_type\": \"\",\n   \"description\": \"\"\n   }\
n   },\n   {\n   \"column\": \"height\",\n   \"properties\":
{\n   \"dtype\": \"number\",\n   \"std\":
0.49497474683058307,\n   \"min\": 1.6,\n   \"max\": 2.3,\n
\"num_unique_values\": 2,\n   \"samples\": [\n   1.6,\n
2.3\n   ],\n   \"semantic_type\": \"\",\n
\"description\": \"\"\n   }\n   }\n]\n}","type":"dataframe"}

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
```

```
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.head(2)
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 4,\n  \"fields\": [\n
{\n      \"column\": \"name\",\n      \"properties\": {\n
\"dtype\": \"string\",\n        \"num_unique_values\": 4,\n
\"samples\": [\n          \"Dave\",\n          \"Charlie\",\n
\"Alice\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"age\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 6,\n        \"min\": 21,\n        \"max\": 34,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          21,\n
34,\n          31\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"gender\",\n      \"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 2,\n        \"samples\": [\n          \"M\",\n
\"F\"\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"height\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 0.39157800414902433,\n        \"min\": 1.6,\n        \"max\":
2.5,\n        \"num_unique_values\": 4,\n        \"samples\": [\n
2.0,\n          1.6\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df"}
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4 entries, 0 to 3
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   name    4 non-null      object
 1   age     4 non-null      int64
 2   gender  4 non-null      object
 3   height  4 non-null      float64
dtypes: float64(1), int64(1), object(2)
memory usage: 256.0+ bytes
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
```

```
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df.describe()
```

```
{"summary":"{\n  \"name\": \"df\",\n  \"rows\": 8,\n  \"fields\": [\n
{\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 11.01086645751157,\n
\"min\": 4.0,\n        \"max\": 34.0,\n        \"num_unique_values\":
8,\n        \"samples\": [\n          27.0,\n          26.5,\n
4.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    },\n    {\n      \"column\":
\"height\",\n      \"properties\": {\n        \"dtype\": \"number\",\n
\"std\": 1.002790246508512,\n        \"min\": 0.39157800414902433,\n
\"max\": 4.0,\n        \"num_unique_values\": 8,\n        \"samples\":
[\n          2.1,\n          2.15,\n          4.0\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    }\n  ]\n}","type":"dataframe"}
```

```
import pandas as pd
data={"name":["Alice","Dave","Bob","Charlie"],
      "age":[31,21,22,34],
      "gender":["F","M","M","F"],
      "height":[2.5,2.0,2.3,1.6]}
df=pd.DataFrame(data)
df_sorted=df.sort_values(by="age",ascending=False)
df_sorted
```

```
{"summary":"{\n  \"name\": \"df_sorted\",\n  \"rows\": 4,\n
\"fields\": [\n    {\n      \"column\": \"name\",\n
\"properties\": {\n        \"dtype\": \"string\",\n
\"num_unique_values\": 4,\n        \"samples\": [\n
\"Alice\",\n          \"Dave\",\n          \"Charlie\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"age\",\n      \"properties\": {\n
\"dtype\": \"number\",\n        \"std\": 6,\n        \"min\": 21,\n
\"max\": 34,\n        \"num_unique_values\": 4,\n        \"samples\":
[\n          31,\n          21,\n          34\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"gender\",\n      \"properties\":
{\n        \"dtype\": \"string\",\n        \"num_unique_values\": 2,\n
\"samples\": [\n          \"M\",\n          \"F\"\n        ],\n
\"semantic_type\": \"\",\n        \"description\": \"\"\n      }\
n    },\n    {\n      \"column\": \"height\",\n      \"properties\":
{\n        \"dtype\": \"number\",\n        \"std\":
0.39157800414902433,\n        \"min\": 1.6,\n        \"max\": 2.5,\n
\"num_unique_values\": 4,\n        \"samples\": [\n          2.5,\n
2.0\n        ],\n        \"semantic_type\": \"\",\n
\"description\": \"\"\n      }\n    }\n  ]\
n}","type":"dataframe","variable_name":"df_sorted"}
```