(http://creativecommons.org/licenses/by-nc-nd/4.0/)
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Spark Tutorial: Learning Apache Spark

This tutorial will teach you how to use Apache Spark (http://spark.apache.org/), a framework for large-scale data processing, within a notebook. Many traditional frameworks were designed to be run on a single computer. However, many datasets today are too large to be stored on a single computer, and even when a dataset can be stored on one computer (such as the datasets in this tutorial), the dataset can often be processed much more quickly using multiple computers.

Spark has efficient implementations of a number of transformations and actions that can be composed together to perform data processing and analysis. Spark excels at distributing these operations across a cluster while abstracting away many of the underlying implementation details. Spark has been designed with a focus on scalability and efficiency. With Spark you can begin developing your solution on your laptop, using a small dataset, and then use that same code to process terabytes or even petabytes across a distributed cluster.

During this tutorial we will cover:

- *Part 1:* Basic notebook usage and Python (https://docs.python.org/2/) integration
- *Part 2:* An introduction to using Apache Spark (https://spark.apache.org/) with the PySpark SQL API (http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#pyspark-sql-module) running in a notebook
- *Part 3:* Using DataFrames and chaining together transformations and actions
- *Part 4*: Python Lambda functions and User Defined Functions
- *Part 5:* Additional DataFrame actions
- *Part 6:* Additional DataFrame transformations
- *Part 7:* Caching DataFrames and storage options
- *Part 8:* Debugging Spark applications and lazy evaluation

The following transformations will be covered:

- `select()`, `filter()`, `distinct()`, `dropDuplicates()`, `orderBy()`, `groupBy()`

The following actions will be covered:

- `first()`, `take()`, `count()`, `collect()`, `show()`

Also covered:

- `cache()`, `unpersist()`

Note that, for reference, you can look up the details of these methods in the Spark's PySpark SQL API (http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#pyspark-sql-module)

Part 1: Basic notebook usage and Python (https://docs.python.org/2/) integration

(1a) Notebook usage

A notebook is comprised of a linear sequence of cells. These cells can contain either markdown or code, but we won't mix both in one cell. When a markdown cell is executed it renders formatted text, images, and links just like HTML in a normal webpage. The text you are reading right now is part of a markdown cell. Python code cells allow you to execute arbitrary Python commands just like in any Python shell. Place your cursor inside the cell below, and press "Shift" + "Enter" to execute the code and advance to the next cell. You can also press "Ctrl" + "Enter" to execute the code and remain in the cell. These commands work the same in both markdown and code cells.

In [2]:
```
# This is a Python cell. You can run normal Python code here...
print 'The sum of 1 and 1 is {0}'.format(1+1)
```

```
The sum of 1 and 1 is 2
```

In [3]:
```
# Here is another Python cell, this time with a variable (x) declaration
 and an if statement:
x = 42
if x > 40:
    print 'The sum of 1 and 2 is {0}'.format(1+2)
```

```
The sum of 1 and 2 is 3
```

(1b) Notebook state

As you work through a notebook it is important that you run all of the code cells. The notebook is stateful, which means that variables and their values are retained until the notebook is detached or the kernel is restarted (in Jupyter notebooks). If you do not run all of the code cells as you proceed through the notebook, your variables will not be properly initialized and later code might fail. You will also need to rerun any cells that you have modified in order for the changes to be available to other cells.

In [4]:
```
# This cell relies on x being defined already.
# If we didn't run the cells from part (1a) this code would fail.
print x * 2
```

```
84
```

(1c) Library imports

We can import standard Python libraries (modules (https://docs.python.org/2/tutorial/modules.html)) the usual way. An `import` statement will import the specified module. In this tutorial and future labs, we will provide any imports that are necessary.

In [5]:
```
# Import the regular expression library
import re
m = re.search('(?<=abc)def', 'abcdef')
m.group(0)
```

Out[5]:
```
def
```

In [6]:
```
# Import the datetime library
import datetime
print 'This was last run on: {0}'.format(datetime.datetime.now())
```

```
This was last run on: 2016-12-08 12:37:05.959206
```

Part 2: An introduction to using Apache Spark (https://spark.apache.org/) with the PySpark SQL API (http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#py sql-module) running in a notebook

Spark Context

In Spark, communication occurs between a driver and executors. The driver has Spark jobs that it needs to run and these jobs are split into tasks that are submitted to the executors for completion. The results from these tasks are delivered back to the driver.

In part 1, we saw that normal Python code can be executed via cells. When using spark-submit this code gets executed in the Spark driver's Java Virtual Machine (JVM) and not in an executor's JVM, and when using a Jupyter notebook it is executed within the kernel associated with the notebook. Since no Spark functionality is actually being used, no tasks are launched on the executors.

In order to use Spark and its DataFrame API we will need to use a `SQLContext`. When running Spark, you start a new Spark application by creating a SparkContext (http://spark.apache.org/docs/latest/api/python/pyspark.html#pyspark.SparkContext). You can then create a SQLContext (http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#pyspark.sql.SQLContext) from the `SparkContext`. When the `SparkContext` is created, it asks the master for some cores to use to do work. The master sets these cores aside just for you; they won't be used for other applications. When using Jupyter, both a `SparkContext` is created for you automatically. `sc` is your `SparkContext`.

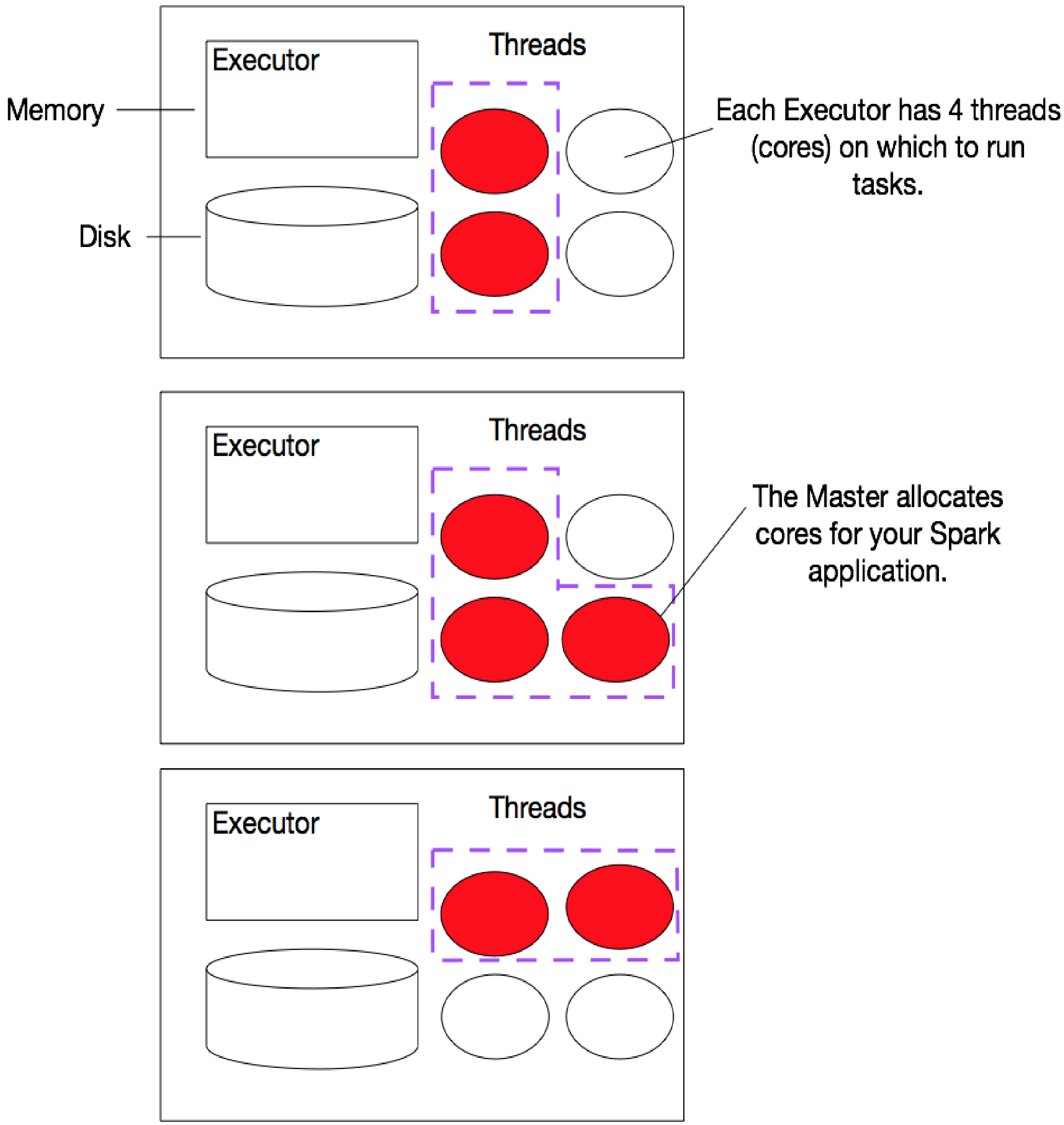
(2a) Example Cluster

The diagram shows an example cluster, where the slots allocated for an application are outlined in purple. (Note: We're using the term *slots* here to indicate threads available to perform parallel work for Spark. Spark documentation often refers to these threads as *cores*, which is a confusing term, as the number of slots available on a particular machine does not necessarily have any relationship to the number of physical CPU cores on that machine.)

You can view the details of your Spark application in the Spark web UI. In the web UI, under the "Jobs" tab, you can see a list of jobs that have been scheduled or run. It's likely there isn't any thing interesting here yet because we haven't run any jobs, but we'll return to this page later.

At a high level, every Spark application consists of a driver program that launches various parallel operations on executor Java Virtual Machines (JVMs) running either in a cluster or locally on the same machine. In Jupyter, "Spark Shell" is the driver program. When running locally, `pyspark` is the driver program. In all cases, this driver program contains the main loop for the program and creates distributed datasets on the cluster, then applies operations (transformations & actions) to those datasets. Driver programs access Spark through a SparkContext object, which represents a connection to a computing cluster. A Spark SQL context object (`sqlContext`) is the main entry point for Spark DataFrame and SQL functionality. A `SQLContext` can be used to create DataFrames, which allows you to direct the operations on your data.

Try printing out `sqlContext` to see its type.



In [7]:
```
# Display the type of the Spark sqlContext
type(sqlContext)
```

Out[7]: <class 'pyspark.sql.context.SQLContext'>

Note that the type is `HiveContext`. This means we're working with a version of Spark that has Hive support. Compiling Spark with Hive support is a good idea, even if you don't have a Hive metastore. As the Spark Programming Guide (http://spark.apache.org/docs/latest/sql-programming-guide.html#starting-point-sqlcontext) states, a `HiveContext` "provides a superset of the functionality provided by the basic `SQLContext`. Additional features include the ability to write queries using the more complete HiveQL parser, access to Hive UDFs [user-defined functions], and the ability to read data from Hive tables. To use a `HiveContext`, you do not need to have an existing Hive setup, and all of the data sources available to a `SQLContext` are still available."

(2b) SparkContext attributes

You can use Python's dir() (https://docs.python.org/2/library/functions.html?highlight=dir#dir) function to get a list of all the attributes (including methods) accessible through the `sqlContext` object.

In [8]:

```
# List sqlContext's attributes
dir(sqlContext)
```

Out[8]:

```
['__class__', '__delattr__', '__dict__', '__doc__', '__format__', '__ge
tattribute__', '__hash__', '__init__', '__module__', '__new__', '__redu
ce__', '__reduce_ex__', '__repr__', '__setattr__', '__sizeof__', '__str
__', '__subclasshook__', '__weakref__', '_createFromLocal', '_createFro
mRDD', '_inferSchema', '_inferSchemaFromList', '_jsc', '_jvm', '_sc',
 '_scala_SQLContext', '_ssql_ctx', 'applySchema', 'cacheTable', 'clearC
ache', 'createDataFrame', 'createExternalTable', 'getConf', 'inferSchem
a', 'jsonFile', 'jsonRDD', 'load', 'parquetFile', 'range', 'read', 'reg
isterDataFrameAsTable', 'registerFunction', 'setConf', 'sql', 'table',
 'tableNames', 'tables', 'udf', 'uncacheTable']
```

(2c) Getting help

Alternatively, you can use Python's help() (https://docs.python.org/2/library/functions.html?highlight=help#help) function to get an easier to read list of all the attributes, including examples, that the `sqlContext` object has.

In [9]:

```
# Use help to obtain more detailed information
help(sqlContext)
```

```
Help on SQLContext in module pyspark.sql.context object:

class SQLContext(__builtin__.object)
 |  Main entry point for Spark SQL functionality.
 |
 |  A SQLContext can be used create :class:`DataFrame`, register :clas
s:`DataFrame` as
 |  tables, execute SQL over tables, cache tables, and read parquet fil
es.
 |
 |  :param sparkContext: The :class:`SparkContext` backing this SQLCont
ext.
 |  :param sqlContext: An optional JVM Scala SQLContext. If set, we do
 not instantiate a new
 |      SQLContext in the JVM, instead we make all calls to this objec
t.
 |
 |  Methods defined here:
 |
 |  __init__(self, sparkContext, sqlContext=None)
 |      Creates a new SQLContext.
 |
 |      >>> from datetime import datetime
 |      >>> sqlContext = SQLContext(sc)
 |      >>> allTypes = sc.parallelize([Row(i=1, s="string", d=1.0, l=1,
 |      ...     b=True, list=[1, 2, 3], dict={"s": 0}, row=Row(a=1),
 |      ...     time=datetime(2014, 8, 1, 14, 1, 5))])
 |      >>> df = allTypes.toDF()
 |      >>> df.registerTempTable("allTypes")
 |      >>> sqlContext.sql('select i+1, d+1, not b, list[1], dict["s"],
time, row.a '
 |      ...            'from allTypes where b and i > 0').collect()
 |      [Row(_c0=2, _c1=2.0, _c2=False, _c3=2, _c4=0,            time=
datetime.datetime(2014, 8, 1, 14, 1, 5), a=1)]
 |      >>> df.map(lambda x: (x.i, x.s, x.d, x.l, x.b, x.time, x.row.a,
x.list)).collect()
 |      [(1, u'string', 1.0, 1, True, datetime.datetime(2014, 8, 1, 14,
1, 5), 1, [1, 2, 3])]
 |
 |  applySchema(self, rdd, schema)
 |      .. note:: Deprecated in 1.3, use :func:`createDataFrame` instea
d.
 |
 |  cacheTable(self, tableName)
 |      Caches the specified table in-memory.
 |
 |      .. versionadded:: 1.0
 |
 |  clearCache(self)
 |      Removes all cached tables from the in-memory cache.
 |
 |      .. versionadded:: 1.3
 |
 |  createDataFrame(self, data, schema=None, samplingRatio=None)
 |      Creates a :class:`DataFrame` from an :class:`RDD` of :class:`tu
ple`/:class:`list`,
 |      list or :class:`pandas.DataFrame`.
```

```
 |
 |      When ``schema`` is a list of column names, the type of each col
umn
 |      will be inferred from ``data``.
 |
 |      When ``schema`` is ``None``, it will try to infer the schema (c
olumn names and types)
 |      from ``data``, which should be an RDD of :class:`Row`,
 |      or :class:`namedtuple`, or :class:`dict`.
 |
 |      If schema inference is needed, ``samplingRatio`` is used to det
ermined the ratio of
 |      rows used for schema inference. The first row will be used if `
`samplingRatio`` is ``None``.
 |
 |      :param data: an RDD of :class:`Row`/:class:`tuple`/:class:`list
`/:class:`dict`,
 |          :class:`list`, or :class:`pandas.DataFrame`.
 |      :param schema: a :class:`StructType` or list of column names. d
efault None.
 |      :param samplingRatio: the sample ratio of rows used for inferri
ng
 |      :return: :class:`DataFrame`
 |
 |      >>> l = [('Alice', 1)]
 |      >>> sqlContext.createDataFrame(l).collect()
 |      [Row(_1=u'Alice', _2=1)]
 |      >>> sqlContext.createDataFrame(l, ['name', 'age']).collect()
 |      [Row(name=u'Alice', age=1)]
 |
 |      >>> d = [{'name': 'Alice', 'age': 1}]
 |      >>> sqlContext.createDataFrame(d).collect()
 |      [Row(age=1, name=u'Alice')]
 |
 |      >>> rdd = sc.parallelize(l)
 |      >>> sqlContext.createDataFrame(rdd).collect()
 |      [Row(_1=u'Alice', _2=1)]
 |      >>> df = sqlContext.createDataFrame(rdd, ['name', 'age'])
 |      >>> df.collect()
 |      [Row(name=u'Alice', age=1)]
 |
 |      >>> from pyspark.sql import Row
 |      >>> Person = Row('name', 'age')
 |      >>> person = rdd.map(lambda r: Person(*r))
 |      >>> df2 = sqlContext.createDataFrame(person)
 |      >>> df2.collect()
 |      [Row(name=u'Alice', age=1)]
 |
 |      >>> from pyspark.sql.types import *
 |      >>> schema = StructType([
 |      ...    StructField("name", StringType(), True),
 |      ...    StructField("age", IntegerType(), True)])
 |      >>> df3 = sqlContext.createDataFrame(rdd, schema)
 |      >>> df3.collect()
 |      [Row(name=u'Alice', age=1)]
 |
 |      >>> sqlContext.createDataFrame(df.toPandas()).collect()  # doct
```

```
est: +SKIP
 |      [Row(name=u'Alice', age=1)]
 |      >>> sqlContext.createDataFrame(pandas.DataFrame([[1, 2]]).colle
ct())  # doctest: +SKIP
 |      [Row(0=1, 1=2)]
 |
 |      .. versionadded:: 1.3
 |
 |  createExternalTable(self, tableName, path=None, source=None, schema
=None, **options)
 |      Creates an external table based on the dataset in a data sourc
e.
 |
 |      It returns the DataFrame associated with the external table.
 |
 |      The data source is specified by the ``source`` and a set of ``o
ptions``.
 |      If ``source`` is not specified, the default data source configu
red by
 |      ``spark.sql.sources.default`` will be used.
 |
 |      Optionally, a schema can be provided as the schema of the retur
ned :class:`DataFrame` and
 |      created external table.
 |
 |      :return: :class:`DataFrame`
 |
 |      .. versionadded:: 1.3
 |
 |  getConf(self, key, defaultValue)
 |      Returns the value of Spark SQL configuration property for the g
iven key.
 |
 |      If the key is not set, returns defaultValue.
 |
 |      .. versionadded:: 1.3
 |
 |  inferSchema(self, rdd, samplingRatio=None)
 |      .. note:: Deprecated in 1.3, use :func:`createDataFrame` instea
d.
 |
 |  jsonFile(self, path, schema=None, samplingRatio=1.0)
 |      Loads a text file storing one JSON object per line as a :class:
`DataFrame`.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameReader.json` i
nstead.
 |
 |      >>> sqlContext.jsonFile('python/test_support/sql/people.json').
dtypes
 |      [('age', 'bigint'), ('name', 'string')]
 |
 |  jsonRDD(self, rdd, schema=None, samplingRatio=1.0)
 |      Loads an RDD storing one JSON object per string as a :class:`Da
taFrame`.
 |
 |      If the schema is provided, applies the given schema to this JSO
```

```
N dataset.
 |      Otherwise, it samples the dataset with ratio ``samplingRatio``
 to determine the schema.
 |
 |      >>> df1 = sqlContext.jsonRDD(json)
 |      >>> df1.first()
 |      Row(field1=1, field2=u'row1', field3=Row(field4=11, field5=Non
e), field6=None)
 |
 |      >>> df2 = sqlContext.jsonRDD(json, df1.schema)
 |      >>> df2.first()
 |      Row(field1=1, field2=u'row1', field3=Row(field4=11, field5=Non
e), field6=None)
 |
 |      >>> from pyspark.sql.types import *
 |      >>> schema = StructType([
 |      ...     StructField("field2", StringType()),
 |      ...     StructField("field3",
 |      ...                 StructType([StructField("field5", ArrayType
(IntegerType()))]))
 |      ... ])
 |      >>> df3 = sqlContext.jsonRDD(json, schema)
 |      >>> df3.first()
 |      Row(field2=u'row1', field3=Row(field5=None))
 |
 |      .. versionadded:: 1.0
 |
 |  load(self, path=None, source=None, schema=None, **options)
 |      Returns the dataset in a data source as a :class:`DataFrame`.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameReader.load` i
nstead.
 |
 |  parquetFile(self, *paths)
 |      Loads a Parquet file, returning the result as a :class:`DataFra
me`.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameReader.parquet
` instead.
 |
 |      >>> sqlContext.parquetFile('python/test_support/sql/parquet_par
titioned').dtypes
 |      [('name', 'string'), ('year', 'int'), ('month', 'int'), ('day',
'int')]
 |
 |  range(self, start, end=None, step=1, numPartitions=None)
 |      Create a :class:`DataFrame` with single LongType column named `
id`,
 |      containing elements in a range from `start` to `end` (exclusiv
e) with
 |      step value `step`.
 |
 |      :param start: the start value
 |      :param end: the end value (exclusive)
 |      :param step: the incremental step (default: 1)
 |      :param numPartitions: the number of partitions of the DataFrame
 |      :return: :class:`DataFrame`
```

```
 |
 |      >>> sqlContext.range(1, 7, 2).collect()
 |      [Row(id=1), Row(id=3), Row(id=5)]
 |
 |      If only one argument is specified, it will be used as the end v
alue.
 |
 |      >>> sqlContext.range(3).collect()
 |      [Row(id=0), Row(id=1), Row(id=2)]
 |
 |      .. versionadded:: 1.4
 |
 |  registerDataFrameAsTable(self, df, tableName)
 |      Registers the given :class:`DataFrame` as a temporary table in
 the catalog.
 |
 |      Temporary tables exist only during the lifetime of this instanc
e of :class:`SQLContext`.
 |
 |      >>> sqlContext.registerDataFrameAsTable(df, "table1")
 |
 |      .. versionadded:: 1.3
 |
 |  registerFunction(self, name, f, returnType=StringType)
 |      Registers a lambda function as a UDF so it can be used in SQL s
tatements.
 |
 |      In addition to a name and the function itself, the return type
 can be optionally specified.
 |      When the return type is not given it default to a string and co
nversion will automatically
 |      be done.  For any other return type, the produced object must m
atch the specified type.
 |
 |      :param name: name of the UDF
 |      :param samplingRatio: lambda function
 |      :param returnType: a :class:`DataType` object
 |
 |      >>> sqlContext.registerFunction("stringLengthString", lambda x:
len(x))
 |      >>> sqlContext.sql("SELECT stringLengthString('test')").collect
()
 |      [Row(_c0=u'4')]
 |
 |      >>> from pyspark.sql.types import IntegerType
 |      >>> sqlContext.registerFunction("stringLengthInt", lambda x: le
n(x), IntegerType())
 |      >>> sqlContext.sql("SELECT stringLengthInt('test')").collect()
 |      [Row(_c0=4)]
 |
 |      >>> from pyspark.sql.types import IntegerType
 |      >>> sqlContext.udf.register("stringLengthInt", lambda x: len
(x), IntegerType())
 |      >>> sqlContext.sql("SELECT stringLengthInt('test')").collect()
 |      [Row(_c0=4)]
 |
 |      .. versionadded:: 1.2
```

```
 |
 |  setConf(self, key, value)
 |      Sets the given Spark SQL configuration property.
 |
 |      .. versionadded:: 1.3
 |
 |  sql(self, sqlQuery)
 |      Returns a :class:`DataFrame` representing the result of the giv
en query.
 |
 |      :return: :class:`DataFrame`
 |
 |      >>> sqlContext.registerDataFrameAsTable(df, "table1")
 |      >>> df2 = sqlContext.sql("SELECT field1 AS f1, field2 as f2 fro
m table1")
 |      >>> df2.collect()
 |      [Row(f1=1, f2=u'row1'), Row(f1=2, f2=u'row2'), Row(f1=3, f2=u'r
ow3')]
 |
 |      .. versionadded:: 1.0
 |
 |  table(self, tableName)
 |      Returns the specified table as a :class:`DataFrame`.
 |
 |      :return: :class:`DataFrame`
 |
 |      >>> sqlContext.registerDataFrameAsTable(df, "table1")
 |      >>> df2 = sqlContext.table("table1")
 |      >>> sorted(df.collect()) == sorted(df2.collect())
 |      True
 |
 |      .. versionadded:: 1.0
 |
 |  tableNames(self, dbName=None)
 |      Returns a list of names of tables in the database ``dbName``.
 |
 |      :param dbName: string, name of the database to use. Default to
 the current database.
 |      :return: list of table names, in string
 |
 |      >>> sqlContext.registerDataFrameAsTable(df, "table1")
 |      >>> "table1" in sqlContext.tableNames()
 |      True
 |      >>> "table1" in sqlContext.tableNames("db")
 |      True
 |
 |      .. versionadded:: 1.3
 |
 |  tables(self, dbName=None)
 |      Returns a :class:`DataFrame` containing names of tables in the
 given database.
 |
 |      If ``dbName`` is not specified, the current database will be us
ed.
 |
 |      The returned DataFrame has two columns: ``tableName`` and ``isT
emporary``
```

```
 |      (a column with :class:`BooleanType` indicating if a table is a
 temporary one or not).
 |
 |      :param dbName: string, name of the database to use.
 |      :return: :class:`DataFrame`
 |
 |      >>> sqlContext.registerDataFrameAsTable(df, "table1")
 |      >>> df2 = sqlContext.tables()
 |      >>> df2.filter("tableName = 'table1'").first()
 |      Row(tableName=u'table1', isTemporary=True)
 |
 |      .. versionadded:: 1.3
 |
 |  uncacheTable(self, tableName)
 |      Removes the specified table from the in-memory cache.
 |
 |      .. versionadded:: 1.0
 |
 |  ----------------------------------------------------------------------
---
 |  Data descriptors defined here:
 |
 |  __dict__
 |      dictionary for instance variables (if defined)
 |
 |  __weakref__
 |      list of weak references to the object (if defined)
 |
 |  read
 |      Returns a :class:`DataFrameReader` that can be used to read dat
a
 |      in as a :class:`DataFrame`.
 |
 |      :return: :class:`DataFrameReader`
 |
 |      .. versionadded:: 1.4
 |
 |  udf
 |      Returns a :class:`UDFRegistration` for UDF registration.
 |
 |      :return: :class:`UDFRegistration`
 |
 |      .. versionadded:: 1.3.1
```

Outside of `pyspark` or a notebook, `SQLContext` is created from the lower-level `SparkContext`, which is usually used to create Resilient Distributed Datasets (RDDs). An RDD is the way Spark actually represents data internally; DataFrames are actually implemented in terms of RDDs.

While you can interact directly with RDDs, DataFrames are preferred. They're generally faster, and they perform the same no matter what language (Python, R, Scala or Java) you use with Spark.

In this course, we'll be using DataFrames, so we won't be interacting directly with the Spark Context object very much. However, it's worth knowing that inside `pyspark` or a notebook, you already have an existing `SparkContext` in the `sc` variable. One simple thing we can do with `sc` is check the version of Spark we're using:

In [10]:
```
# After reading the help we've decided we want to use sc.version to see
what version of Spark we are running
sc.version
```

Out[10]:
```
1.5.2
```

In [11]:
```
# Help can be used on any Python object
help(map)
```

```
Help on built-in function map in module __builtin__:

map(...)
    map(function, sequence[, sequence, ...]) -> list

    Return a list of the results of applying the function to the items
 of
    the argument sequence(s).  If more than one sequence is given, the
    function is called with an argument list consisting of the correspo
nding
    item of each sequence, substituting None for missing values when no
t all
    sequences have the same length.  If the function is None, return a
 list of
    the items of the sequence (or a list of tuples if more than one seq
uence).
```

Part 3: Using DataFrames and chaining together transformations and actions

Working with your first DataFrames

In Spark, we first create a base DataFrame (http://spark.apache.org/docs/latest/api/python/pyspark.sql.html#pyspark.sql.DataFrame). We can then apply one or more transformations to that base DataFrame. *A DataFrame is immutable, so once it is created, it cannot be changed.* As a result, each transformation creates a new DataFrame. Finally, we can apply one or more actions to the DataFrames.

> Note that Spark uses lazy evaluation, so transformations are not actually executed until an action occurs.

We will perform several exercises to obtain a better understanding of DataFrames:

- Create a Python collection of 10,000 integers
- Create a Spark DataFrame from that collection
- Subtract one from each value using `map`
- Perform action `collect` to view results
- Perform action `count` to view counts
- Apply transformation `filter` and view results with `collect`
- Learn about lambda functions
- Explore how lazy evaluation works and the debugging challenges that it introduces

A DataFrame consists of a series of `Row` objects; each `Row` object has a set of named columns. You can think of a DataFrame as modeling a table, though the data source being processed does not have to be a table.

More formally, a DataFrame must have a *schema*, which means it must consist of columns, each of which has a *name* and a *type*. Some data sources have schemas built into them. Examples include RDBMS databases, Parquet files, and NoSQL databases like Cassandra. Other data sources don't have computer-readable schemas, but you can often apply a schema programmatically.

(3a) Create a Python collection of 10,000 people

We will use a third-party Python testing library called fake-factory (https://pypi.python.org/pypi/fake-factory/0.5.3) to create a collection of fake person records.

From the command line on login1 node, run the following command and restart this kernel:

```
pip install --upgrade fake-factory --user
```

In [12]:
```
from faker import Factory
fake = Factory.create()
fake.seed(4321)
```

We're going to use this factory to create a collection of randomly generated people records. In the next section, we'll turn that collection into a DataFrame. We'll use the Spark `Row` class, because that will help us define the Spark DataFrame schema. There are other ways to define schemas, though; see the Spark Programming Guide's discussion of [schema inference (http://spark.apache.org/docs/latest/sql-programming-guide.html#inferring-the-schema-using-reflection)](http://spark.apache.org/docs/latest/sql-programming-guide.html#inferring-the-schema-using-reflection) for more information. (For instance, we could also use a Python `namedtuple`.)

In [13]:
```
# Each entry consists of last_name, first_name, ssn, job, and age (at le
ast 1)
from pyspark.sql import Row
def fake_entry():
  name = fake.name().split()
  return (name[1], name[0], fake.ssn(), fake.job(), abs(2016 - fake.date
_time().year) + 1)
```

In [14]:
```
# Create a helper function to call a function repeatedly
def repeat(times, func, *args, **kwargs):
    for _ in xrange(times):
        yield func(*args, **kwargs)
```

In [15]:
```
data = list(repeat(10000, fake_entry))
```

`data` is just a normal Python list, containing Python tuples objects. Let's look at the first item in the list:

In [16]:
```
data[0]
```

Out[16]:
```
(u'Brown', u'Jason', u'160-37-9051', 'Agricultural engineer', 39)
```

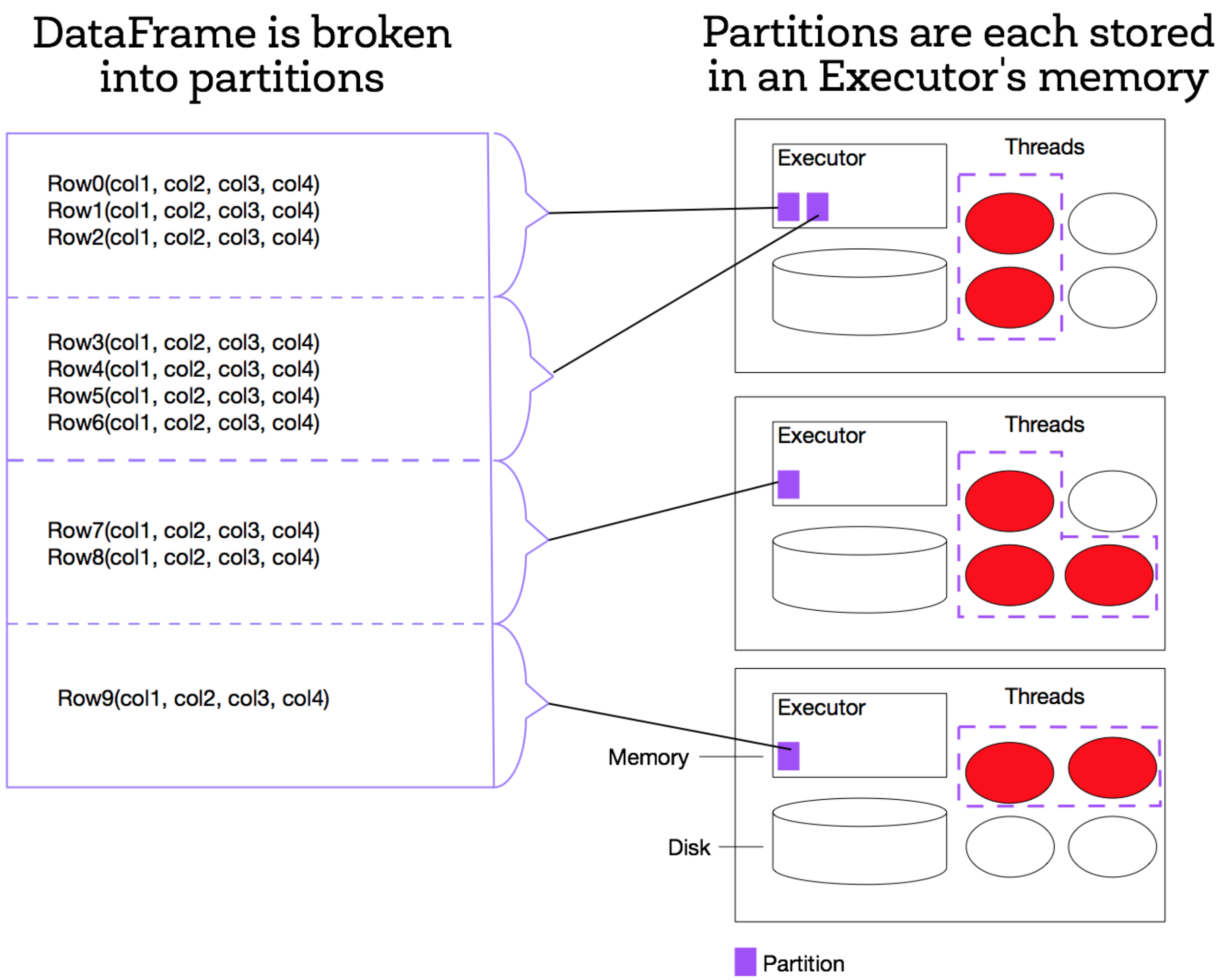
We can check the size of the list using the Python `len()` function.

In [17]:
```
len(data)
```

Out[17]:
```
10000
```

(3b) Distributed data and using a collection to create a DataFrame

In Spark, datasets are represented as a list of entries, where the list is broken up into many different partitions that are each stored on a different machine. Each partition holds a unique subset of the entries in the list. Spark calls datasets that it stores "Resilient Distributed Datasets" (RDDs). Even DataFrames are ultimately represented as RDDs, with additional meta-data.



One of the defining features of Spark, compared to other data analytics frameworks (e.g., Hadoop), is that it stores data in memory rather than on disk. This allows Spark applications to run much more quickly, because they are not slowed down by needing to read data from disk. The figure to the right illustrates how Spark breaks a list of data entries into partitions that are each stored in memory on a worker.

To create the DataFrame, we'll use `sqlContext.createDataFrame()`, and we'll pass our array of data in as an argument to that function. Spark will create a new set of input data based on data that is passed in. A DataFrame requires a *schema*, which is a list of columns, where each column has a name and a type. Our list of data has elements with types (mostly strings, but one integer). We'll supply the rest of the schema and the column names as the second argument to `createDataFrame()`.

Let's view the help for `createDataFrame()`.

In [18]:
```
help(sqlContext.createDataFrame)
```

```
Help on method createDataFrame in module pyspark.sql.context:

createDataFrame(self, data, schema=None, samplingRatio=None) method of
 pyspark.sql.context.SQLContext instance
    Creates a :class:`DataFrame` from an :class:`RDD` of :class:`tuple
`/:class:`list`,
    list or :class:`pandas.DataFrame`.

    When ``schema`` is a list of column names, the type of each column
    will be inferred from ``data``.

    When ``schema`` is ``None``, it will try to infer the schema (colum
n names and types)
    from ``data``, which should be an RDD of :class:`Row`,
    or :class:`namedtuple`, or :class:`dict`.

    If schema inference is needed, ``samplingRatio`` is used to determi
ned the ratio of
    rows used for schema inference. The first row will be used if ``sam
plingRatio`` is ``None``.

    :param data: an RDD of :class:`Row`/:class:`tuple`/:class:`list`/:c
lass:`dict`,
        :class:`list`, or :class:`pandas.DataFrame`.
    :param schema: a :class:`StructType` or list of column names. defau
lt None.
    :param samplingRatio: the sample ratio of rows used for inferring
    :return: :class:`DataFrame`

    >>> l = [('Alice', 1)]
    >>> sqlContext.createDataFrame(l).collect()
    [Row(_1=u'Alice', _2=1)]
    >>> sqlContext.createDataFrame(l, ['name', 'age']).collect()
    [Row(name=u'Alice', age=1)]

    >>> d = [{'name': 'Alice', 'age': 1}]
    >>> sqlContext.createDataFrame(d).collect()
    [Row(age=1, name=u'Alice')]

    >>> rdd = sc.parallelize(l)
    >>> sqlContext.createDataFrame(rdd).collect()
    [Row(_1=u'Alice', _2=1)]
    >>> df = sqlContext.createDataFrame(rdd, ['name', 'age'])
    >>> df.collect()
    [Row(name=u'Alice', age=1)]

    >>> from pyspark.sql import Row
    >>> Person = Row('name', 'age')
    >>> person = rdd.map(lambda r: Person(*r))
    >>> df2 = sqlContext.createDataFrame(person)
    >>> df2.collect()
    [Row(name=u'Alice', age=1)]

    >>> from pyspark.sql.types import *
    >>> schema = StructType([
    ...    StructField("name", StringType(), True),
    ...    StructField("age", IntegerType(), True)])
```

```
    >>> df3 = sqlContext.createDataFrame(rdd, schema)
    >>> df3.collect()
    [Row(name=u'Alice', age=1)]

    >>> sqlContext.createDataFrame(df.toPandas()).collect()  # doctest:
+SKIP
    [Row(name=u'Alice', age=1)]
    >>> sqlContext.createDataFrame(pandas.DataFrame([[1, 2]]).collect
())  # doctest: +SKIP
    [Row(0=1, 1=2)]

    .. versionadded:: 1.3
```

In [19]:
```
dataDF = sqlContext.createDataFrame(data, ('last_name', 'first_name', 'ssn', 'occupation', 'age'))
```

Let's see what type `sqlContext.createDataFrame()` returned.

In [20]:
```
print 'type of dataDF: {0}'.format(type(dataDF))
```

```
type of dataDF: <class 'pyspark.sql.dataframe.DataFrame'>
```

Let's take a look at the DataFrame's schema and some of its rows.

In [21]:
```
dataDF.printSchema()
```

```
root
 |-- last_name: string (nullable = true)
 |-- first_name: string (nullable = true)
 |-- ssn: string (nullable = true)
 |-- occupation: string (nullable = true)
 |-- age: long (nullable = true)
```

We can register the newly created DataFrame as a named table, using the `registerDataFrameAsTable()` method.

In [22]:
```
sqlContext.registerDataFrameAsTable(dataDF, 'dataframe')
```

What methods can we call on this DataFrame?

In [23]:
```
help(dataDF)
```

```
Help on DataFrame in module pyspark.sql.dataframe object:

class DataFrame(__builtin__.object)
 |  A distributed collection of data grouped into named columns.
 |
 |  A :class:`DataFrame` is equivalent to a relational table in Spark S
QL,
 |  and can be created using various functions in :class:`SQLContext`::
 |
 |      people = sqlContext.read.parquet("...")
 |
 |  Once created, it can be manipulated using the various domain-specif
ic-language
 |  (DSL) functions defined in: :class:`DataFrame`, :class:`Column`.
 |
 |  To select a column from the data frame, use the apply method::
 |
 |      ageCol = people.age
 |
 |  A more concrete example::
 |
 |      # To create DataFrame using SQLContext
 |      people = sqlContext.read.parquet("...")
 |      department = sqlContext.read.parquet("...")
 |
 |      people.filter(people.age > 30).join(department, people.deptId =
= department.id))           .groupBy(department.name, "gender").agg({"s
alary": "avg", "age": "max"})
 |
 |  .. note:: Experimental
 |
 |  .. versionadded:: 1.3
 |
 |  Methods defined here:
 |
 |  __getattr__(self, name)
 |      Returns the :class:`Column` denoted by ``name``.
 |
 |      >>> df.select(df.age).collect()
 |      [Row(age=2), Row(age=5)]
 |
 |      .. versionadded:: 1.3
 |
 |  __getitem__(self, item)
 |      Returns the column as a :class:`Column`.
 |
 |      >>> df.select(df['age']).collect()
 |      [Row(age=2), Row(age=5)]
 |      >>> df[ ["name", "age"]].collect()
 |      [Row(name=u'Alice', age=2), Row(name=u'Bob', age=5)]
 |      >>> df[ df.age > 3 ].collect()
 |      [Row(age=5, name=u'Bob')]
 |      >>> df[df[0] > 3].collect()
 |      [Row(age=5, name=u'Bob')]
 |
 |      .. versionadded:: 1.3
 |
```

```
 |  __init__(self, jdf, sql_ctx)
 |
 |  __repr__(self)
 |
 |  agg(self, *exprs)
 |      Aggregate on the entire :class:`DataFrame` without groups
 |      (shorthand for ``df.groupBy.agg()``).
 |
 |      >>> df.agg({"age": "max"}).collect()
 |      [Row(max(age)=5)]
 |      >>> from pyspark.sql import functions as F
 |      >>> df.agg(F.min(df.age)).collect()
 |      [Row(min(age)=2)]
 |
 |      .. versionadded:: 1.3
 |
 |  alias(self, alias)
 |      Returns a new :class:`DataFrame` with an alias set.
 |
 |      >>> from pyspark.sql.functions import *
 |      >>> df_as1 = df.alias("df_as1")
 |      >>> df_as2 = df.alias("df_as2")
 |      >>> joined_df = df_as1.join(df_as2, col("df_as1.name") == col
("df_as2.name"), 'inner')
 |      >>> joined_df.select(col("df_as1.name"), col("df_as2.name"), co
l("df_as2.age")).collect()
 |      [Row(name=u'Alice', name=u'Alice', age=2), Row(name=u'Bob', nam
e=u'Bob', age=5)]
 |
 |      .. versionadded:: 1.3
 |
 |  cache(self)
 |      Persists with the default storage level (C{MEMORY_ONLY_SER}).
 |
 |      .. versionadded:: 1.3
 |
 |  coalesce(self, numPartitions)
 |      Returns a new :class:`DataFrame` that has exactly `numPartition
s` partitions.
 |
 |      Similar to coalesce defined on an :class:`RDD`, this operation
 results in a
 |      narrow dependency, e.g. if you go from 1000 partitions to 100 p
artitions,
 |      there will not be a shuffle, instead each of the 100 new partit
ions will
 |      claim 10 of the current partitions.
 |
 |      >>> df.coalesce(1).rdd.getNumPartitions()
 |      1
 |
 |      .. versionadded:: 1.4
 |
 |  collect(self)
 |      Returns all the records as a list of :class:`Row`.
 |
 |      >>> df.collect()
```

```
 |      [Row(age=2, name=u'Alice'), Row(age=5, name=u'Bob')]
 |
 |      .. versionadded:: 1.3
 |
 |  corr(self, col1, col2, method=None)
 |      Calculates the correlation of two columns of a DataFrame as a d
ouble value.
 |      Currently only supports the Pearson Correlation Coefficient.
 |      :func:`DataFrame.corr` and :func:`DataFrameStatFunctions.corr`
 are aliases of each other.
 |
 |      :param col1: The name of the first column
 |      :param col2: The name of the second column
 |      :param method: The correlation method. Currently only supports
 "pearson"
 |
 |      .. versionadded:: 1.4
 |
 |  count(self)
 |      Returns the number of rows in this :class:`DataFrame`.
 |
 |      >>> df.count()
 |      2
 |
 |      .. versionadded:: 1.3
 |
 |  cov(self, col1, col2)
 |      Calculate the sample covariance for the given columns, specifie
d by their names, as a
 |      double value. :func:`DataFrame.cov` and :func:`DataFrameStatFun
ctions.cov` are aliases.
 |
 |      :param col1: The name of the first column
 |      :param col2: The name of the second column
 |
 |      .. versionadded:: 1.4
 |
 |  crosstab(self, col1, col2)
 |      Computes a pair-wise frequency table of the given columns. Also
known as a contingency
 |      table. The number of distinct values for each column should be
 less than 1e4. At most 1e6
 |      non-zero pair frequencies will be returned.
 |      The first column of each row will be the distinct values of `co
l1` and the column names
 |      will be the distinct values of `col2`. The name of the first co
lumn will be `$col1_$col2`.
 |      Pairs that have no occurrences will have zero as their counts.
 |      :func:`DataFrame.crosstab` and :func:`DataFrameStatFunctions.cr
osstab` are aliases.
 |
 |      :param col1: The name of the first column. Distinct items will
 make the first item of
 |          each row.
 |      :param col2: The name of the second column. Distinct items will
make the column names
 |          of the DataFrame.
```

```
 |
 |      .. versionadded:: 1.4
 |
 |  cube(self, *cols)
 |      Create a multi-dimensional cube for the current :class:`DataFra
me` using
 |      the specified columns, so we can run aggregation on them.
 |
 |      >>> df.cube('name', df.age).count().show()
 |      +-----+----+-----+
 |      | name| age|count|
 |      +-----+----+-----+
 |      | null|   2|    1|
 |      |Alice|null|    1|
 |      |  Bob|   5|    1|
 |      |  Bob|null|    1|
 |      | null|   5|    1|
 |      | null|null|    2|
 |      |Alice|   2|    1|
 |      +-----+----+-----+
 |
 |      .. versionadded:: 1.4
 |
 |  describe(self, *cols)
 |      Computes statistics for numeric columns.
 |
 |      This include count, mean, stddev, min, and max. If no columns a
re
 |      given, this function computes statistics for all numerical colu
mns.
 |
 |      .. note:: This function is meant for exploratory data analysis,
as we make no         guarantee about the backward compatibility of the
schema of the resulting DataFrame.
 |
 |      >>> df.describe().show()
 |      +-------+---+
 |      |summary|age|
 |      +-------+---+
 |      |  count|  2|
 |      |   mean|3.5|
 |      | stddev|1.5|
 |      |    min|  2|
 |      |    max|  5|
 |      +-------+---+
 |      >>> df.describe(['age', 'name']).show()
 |      +-------+---+-----+
 |      |summary|age| name|
 |      +-------+---+-----+
 |      |  count|  2|    2|
 |      |   mean|3.5| null|
 |      | stddev|1.5| null|
 |      |    min|  2|Alice|
 |      |    max|  5|  Bob|
 |      +-------+---+-----+
 |
 |      .. versionadded:: 1.3.1
```

```
 |
 |  distinct(self)
 |      Returns a new :class:`DataFrame` containing the distinct rows i
n this :class:`DataFrame`.
 |
 |      >>> df.distinct().count()
 |      2
 |
 |      .. versionadded:: 1.3
 |
 |  drop(self, col)
 |      Returns a new :class:`DataFrame` that drops the specified colum
n.
 |
 |      :param col: a string name of the column to drop, or a
 |          :class:`Column` to drop.
 |
 |      >>> df.drop('age').collect()
 |      [Row(name=u'Alice'), Row(name=u'Bob')]
 |
 |      >>> df.drop(df.age).collect()
 |      [Row(name=u'Alice'), Row(name=u'Bob')]
 |
 |      >>> df.join(df2, df.name == df2.name, 'inner').drop(df.name).co
llect()
 |      [Row(age=5, height=85, name=u'Bob')]
 |
 |      >>> df.join(df2, df.name == df2.name, 'inner').drop(df2.name).c
ollect()
 |      [Row(age=5, name=u'Bob', height=85)]
 |
 |      .. versionadded:: 1.4
 |
 |  dropDuplicates(self, subset=None)
 |      Return a new :class:`DataFrame` with duplicate rows removed,
 |      optionally only considering certain columns.
 |
 |      >>> from pyspark.sql import Row
 |      >>> df = sc.parallelize([             Row(name='Alice', age=5,
 height=80),             Row(name='Alice', age=5, height=80),
   Row(name='Alice', age=10, height=80)]).toDF()
 |      >>> df.dropDuplicates().show()
 |      +---+------+-----+
 |      |age|height| name|
 |      +---+------+-----+
 |      |  5|    80|Alice|
 |      | 10|    80|Alice|
 |      +---+------+-----+
 |
 |      >>> df.dropDuplicates(['name', 'height']).show()
 |      +---+------+-----+
 |      |age|height| name|
 |      +---+------+-----+
 |      |  5|    80|Alice|
 |      +---+------+-----+
 |
 |      .. versionadded:: 1.4
```

```
 |
 |  drop_duplicates = dropDuplicates(self, subset=None)
 |
 |  dropna(self, how='any', thresh=None, subset=None)
 |      Returns a new :class:`DataFrame` omitting rows with null value
s.
 |      :func:`DataFrame.dropna` and :func:`DataFrameNaFunctions.drop`
 are aliases of each other.
 |
 |      :param how: 'any' or 'all'.
 |          If 'any', drop a row if it contains any nulls.
 |          If 'all', drop a row only if all its values are null.
 |      :param thresh: int, default None
 |          If specified, drop rows that have less than `thresh` non-nu
ll values.
 |          This overwrites the `how` parameter.
 |      :param subset: optional list of column names to consider.
 |
 |      >>> df4.na.drop().show()
 |      +---+------+-----+
 |      |age|height| name|
 |      +---+------+-----+
 |      | 10|    80|Alice|
 |      +---+------+-----+
 |
 |      .. versionadded:: 1.3.1
 |
 |  explain(self, extended=False)
 |      Prints the (logical and physical) plans to the console for debu
gging purpose.
 |
 |      :param extended: boolean, default ``False``. If ``False``, prin
ts only the physical plan.
 |
 |      >>> df.explain()
 |      Scan PhysicalRDD[age#0,name#1]
 |
 |      >>> df.explain(True)
 |      == Parsed Logical Plan ==
 |      ...
 |      == Analyzed Logical Plan ==
 |      ...
 |      == Optimized Logical Plan ==
 |      ...
 |      == Physical Plan ==
 |      ...
 |
 |      .. versionadded:: 1.3
 |
 |  fillna(self, value, subset=None)
 |      Replace null values, alias for ``na.fill()``.
 |      :func:`DataFrame.fillna` and :func:`DataFrameNaFunctions.fill`
 are aliases of each other.
 |
 |      :param value: int, long, float, string, or dict.
 |          Value to replace null values with.
 |          If the value is a dict, then `subset` is ignored and `value
```

```
` must be a mapping
 |          from column name (string) to replacement value. The replace
ment value must be
 |          an int, long, float, or string.
 |      :param subset: optional list of column names to consider.
 |          Columns specified in subset that do not have matching data
 type are ignored.
 |          For example, if `value` is a string, and subset contains a
 non-string column,
 |          then the non-string column is simply ignored.
 |
 |      >>> df4.na.fill(50).show()
 |      +---+------+-----+
 |      |age|height| name|
 |      +---+------+-----+
 |      | 10|    80|Alice|
 |      |  5|    50|  Bob|
 |      | 50|    50|  Tom|
 |      | 50|    50| null|
 |      +---+------+-----+
 |
 |      >>> df4.na.fill({'age': 50, 'name': 'unknown'}).show()
 |      +---+------+-------+
 |      |age|height|   name|
 |      +---+------+-------+
 |      | 10|    80|  Alice|
 |      |  5|  null|    Bob|
 |      | 50|  null|    Tom|
 |      | 50|  null|unknown|
 |      +---+------+-------+
 |
 |      .. versionadded:: 1.3.1
 |
 |  filter(self, condition)
 |      Filters rows using the given condition.
 |
 |      :func:`where` is an alias for :func:`filter`.
 |
 |      :param condition: a :class:`Column` of :class:`types.BooleanTyp
e`
 |          or a string of SQL expression.
 |
 |      >>> df.filter(df.age > 3).collect()
 |      [Row(age=5, name=u'Bob')]
 |      >>> df.where(df.age == 2).collect()
 |      [Row(age=2, name=u'Alice')]
 |
 |      >>> df.filter("age > 3").collect()
 |      [Row(age=5, name=u'Bob')]
 |      >>> df.where("age = 2").collect()
 |      [Row(age=2, name=u'Alice')]
 |
 |      .. versionadded:: 1.3
 |
 |  first(self)
 |      Returns the first row as a :class:`Row`.
 |
```

```
 |      >>> df.first()
 |      Row(age=2, name=u'Alice')
 |
 |      .. versionadded:: 1.3
 |
 |  flatMap(self, f)
 |      Returns a new :class:`RDD` by first applying the ``f`` function
to each :class:`Row`,
 |      and then flattening the results.
 |
 |      This is a shorthand for ``df.rdd.flatMap()``.
 |
 |      >>> df.flatMap(lambda p: p.name).collect()
 |      [u'A', u'l', u'i', u'c', u'e', u'B', u'o', u'b']
 |
 |      .. versionadded:: 1.3
 |
 |  foreach(self, f)
 |      Applies the ``f`` function to all :class:`Row` of this :class:`
DataFrame`.
 |
 |      This is a shorthand for ``df.rdd.foreach()``.
 |
 |      >>> def f(person):
 |      ...     print(person.name)
 |      >>> df.foreach(f)
 |
 |      .. versionadded:: 1.3
 |
 |  foreachPartition(self, f)
 |      Applies the ``f`` function to each partition of this :class:`Da
taFrame`.
 |
 |      This a shorthand for ``df.rdd.foreachPartition()``.
 |
 |      >>> def f(people):
 |      ...     for person in people:
 |      ...         print(person.name)
 |      >>> df.foreachPartition(f)
 |
 |      .. versionadded:: 1.3
 |
 |  freqItems(self, cols, support=None)
 |      Finding frequent items for columns, possibly with false positiv
es. Using the
 |      frequent element count algorithm described in
 |      "http://dx.doi.org/10.1145/762471.762473, proposed by Karp, Sch
enker, and Papadimitriou".
 |      :func:`DataFrame.freqItems` and :func:`DataFrameStatFunctions.f
reqItems` are aliases.
 |
 |      .. note::  This function is meant for exploratory data analysi
s, as we make no         guarantee about the backward compatibility of
 the schema of the resulting DataFrame.
 |
 |      :param cols: Names of the columns to calculate frequent items f
or as a list or tuple of
```

```
 |          strings.
 |      :param support: The frequency with which to consider an item 'f
requent'. Default is 1%.
 |          The support must be greater than 1e-4.
 |
 |      .. versionadded:: 1.4
 |
 |  groupBy(self, *cols)
 |      Groups the :class:`DataFrame` using the specified columns,
 |      so we can run aggregation on them. See :class:`GroupedData`
 |      for all the available aggregate functions.
 |
 |      :func:`groupby` is an alias for :func:`groupBy`.
 |
 |      :param cols: list of columns to group by.
 |          Each element should be a column name (string) or an express
ion (:class:`Column`).
 |
 |      >>> df.groupBy().avg().collect()
 |      [Row(avg(age)=3.5)]
 |      >>> df.groupBy('name').agg({'age': 'mean'}).collect()
 |      [Row(name=u'Alice', avg(age)=2.0), Row(name=u'Bob', avg(age)=5.
0)]
 |      >>> df.groupBy(df.name).avg().collect()
 |      [Row(name=u'Alice', avg(age)=2.0), Row(name=u'Bob', avg(age)=5.
0)]
 |      >>> df.groupBy(['name', df.age]).count().collect()
 |      [Row(name=u'Bob', age=5, count=1), Row(name=u'Alice', age=2, co
unt=1)]
 |
 |      .. versionadded:: 1.3
 |
 |  groupby = groupBy(self, *cols)
 |
 |  head(self, n=None)
 |      Returns the first ``n`` rows.
 |
 |      :param n: int, default 1. Number of rows to return.
 |      :return: If n is greater than 1, return a list of :class:`Row`.
 |          If n is 1, return a single Row.
 |
 |      >>> df.head()
 |      Row(age=2, name=u'Alice')
 |      >>> df.head(1)
 |      [Row(age=2, name=u'Alice')]
 |
 |      .. versionadded:: 1.3
 |
 |  insertInto(self, tableName, overwrite=False)
 |      Inserts the contents of this :class:`DataFrame` into the specif
ied table.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameWriter.insertI
nto` instead.
 |
 |  intersect(self, other)
 |      Return a new :class:`DataFrame` containing rows only in
```

```
 |      both this frame and another frame.
 |
 |      This is equivalent to `INTERSECT` in SQL.
 |
 |      .. versionadded:: 1.3
 |
 |  isLocal(self)
 |      Returns ``True`` if the :func:`collect` and :func:`take` method
s can be run locally
 |      (without any Spark executors).
 |
 |      .. versionadded:: 1.3
 |
 |  join(self, other, on=None, how=None)
 |      Joins with another :class:`DataFrame`, using the given join exp
ression.
 |
 |      The following performs a full outer join between ``df1`` and ``
df2``.
 |
 |      :param other: Right side of the join
 |      :param on: a string for join column name, a list of column name
s,
 |          , a join expression (Column) or a list of Columns.
 |          If `on` is a string or a list of string indicating the name
of the join column(s),
 |          the column(s) must exist on both sides, and this performs a
n inner equi-join.
 |      :param how: str, default 'inner'.
 |          One of `inner`, `outer`, `left_outer`, `right_outer`, `semi
join`.
 |
 |      >>> df.join(df2, df.name == df2.name, 'outer').select(df.name,
 df2.height).collect()
 |      [Row(name=None, height=80), Row(name=u'Alice', height=None), Ro
w(name=u'Bob', height=85)]
 |
 |      >>> cond = [df.name == df3.name, df.age == df3.age]
 |      >>> df.join(df3, cond, 'outer').select(df.name, df3.age).collec
t()
 |      [Row(name=u'Bob', age=5), Row(name=u'Alice', age=2)]
 |
 |      >>> df.join(df2, 'name').select(df.name, df2.height).collect()
 |      [Row(name=u'Bob', height=85)]
 |
 |      >>> df.join(df4, ['name', 'age']).select(df.name, df.age).colle
ct()
 |      [Row(name=u'Bob', age=5)]
 |
 |      .. versionadded:: 1.3
 |
 |  limit(self, num)
 |      Limits the result count to the number specified.
 |
 |      >>> df.limit(1).collect()
 |      [Row(age=2, name=u'Alice')]
 |      >>> df.limit(0).collect()
```

```
 |      []
 |
 |      .. versionadded:: 1.3
 |
 |  map(self, f)
 |      Returns a new :class:`RDD` by applying a the ``f`` function to
 each :class:`Row`.
 |
 |      This is a shorthand for ``df.rdd.map()``.
 |
 |      >>> df.map(lambda p: p.name).collect()
 |      [u'Alice', u'Bob']
 |
 |      .. versionadded:: 1.3
 |
 |  mapPartitions(self, f, preservesPartitioning=False)
 |      Returns a new :class:`RDD` by applying the ``f`` function to ea
ch partition.
 |
 |      This is a shorthand for ``df.rdd.mapPartitions()``.
 |
 |      >>> rdd = sc.parallelize([1, 2, 3, 4], 4)
 |      >>> def f(iterator): yield 1
 |      >>> rdd.mapPartitions(f).sum()
 |      4
 |
 |      .. versionadded:: 1.3
 |
 |  orderBy = sort(self, *cols, **kwargs)
 |
 |  persist(self, storageLevel=StorageLevel(False, True, False, False,
 1))
 |      Sets the storage level to persist its values across operations
 |      after the first time it is computed. This can only be used to a
ssign
 |      a new storage level if the RDD does not have a storage level se
t yet.
 |      If no storage level is specified defaults to (C{MEMORY_ONLY_SE
R}).
 |
 |      .. versionadded:: 1.3
 |
 |  printSchema(self)
 |      Prints out the schema in the tree format.
 |
 |      >>> df.printSchema()
 |      root
 |       |-- age: integer (nullable = true)
 |       |-- name: string (nullable = true)
 |      <BLANKLINE>
 |
 |      .. versionadded:: 1.3
 |
 |  randomSplit(self, weights, seed=None)
 |      Randomly splits this :class:`DataFrame` with the provided weigh
ts.
 |
```

```
 |      :param weights: list of doubles as weights with which to split
 the DataFrame. Weights will
 |          be normalized if they don't sum up to 1.0.
 |      :param seed: The seed for sampling.
 |
 |      >>> splits = df4.randomSplit([1.0, 2.0], 24)
 |      >>> splits[0].count()
 |      1
 |
 |      >>> splits[1].count()
 |      3
 |
 |      .. versionadded:: 1.4
 |
 |  registerAsTable(self, name)
 |      .. note:: Deprecated in 1.4, use :func:`registerTempTable` inst
ead.
 |
 |  registerTempTable(self, name)
 |      Registers this RDD as a temporary table using the given name.
 |
 |      The lifetime of this temporary table is tied to the :class:`SQL
Context`
 |      that was used to create this :class:`DataFrame`.
 |
 |      >>> df.registerTempTable("people")
 |      >>> df2 = sqlContext.sql("select * from people")
 |      >>> sorted(df.collect()) == sorted(df2.collect())
 |      True
 |
 |      .. versionadded:: 1.3
 |
 |  repartition(self, numPartitions)
 |      Returns a new :class:`DataFrame` that has exactly ``numPartitio
ns`` partitions.
 |
 |      >>> df.repartition(10).rdd.getNumPartitions()
 |      10
 |
 |      .. versionadded:: 1.3
 |
 |  replace(self, to_replace, value, subset=None)
 |      Returns a new :class:`DataFrame` replacing a value with another
value.
 |      :func:`DataFrame.replace` and :func:`DataFrameNaFunctions.repla
ce` are
 |      aliases of each other.
 |
 |      :param to_replace: int, long, float, string, or list.
 |          Value to be replaced.
 |          If the value is a dict, then `value` is ignored and `to_rep
lace` must be a
 |          mapping from column name (string) to replacement value. The
value to be
 |          replaced must be an int, long, float, or string.
 |      :param value: int, long, float, string, or list.
 |          Value to use to replace holes.
```

```
 |          The replacement value must be an int, long, float, or strin
g. If `value` is a
 |          list or tuple, `value` should be of the same length with `t
o_replace`.
 |      :param subset: optional list of column names to consider.
 |          Columns specified in subset that do not have matching data
 type are ignored.
 |          For example, if `value` is a string, and subset contains a
 non-string column,
 |          then the non-string column is simply ignored.
 |
 |      >>> df4.na.replace(10, 20).show()
 |      +----+------+-----+
 |      | age|height| name|
 |      +----+------+-----+
 |      |  20|    80|Alice|
 |      |   5|  null|  Bob|
 |      |null|  null|  Tom|
 |      |null|  null| null|
 |      +----+------+-----+
 |
 |      >>> df4.na.replace(['Alice', 'Bob'], ['A', 'B'], 'name').show()
 |      +----+------+----+
 |      | age|height|name|
 |      +----+------+----+
 |      |  10|    80|   A|
 |      |   5|  null|   B|
 |      |null|  null| Tom|
 |      |null|  null|null|
 |      +----+------+----+
 |
 |      .. versionadded:: 1.4
 |
 |  rollup(self, *cols)
 |      Create a multi-dimensional rollup for the current :class:`DataF
rame` using
 |      the specified columns, so we can run aggregation on them.
 |
 |      >>> df.rollup('name', df.age).count().show()
 |      +-----+----+-----+
 |      | name| age|count|
 |      +-----+----+-----+
 |      |Alice|null|    1|
 |      |  Bob|   5|    1|
 |      |  Bob|null|    1|
 |      | null|null|    2|
 |      |Alice|   2|    1|
 |      +-----+----+-----+
 |
 |      .. versionadded:: 1.4
 |
 |  sample(self, withReplacement, fraction, seed=None)
 |      Returns a sampled subset of this :class:`DataFrame`.
 |
 |      >>> df.sample(False, 0.5, 42).count()
 |      1
 |
```

```
 |      .. versionadded:: 1.3
 |
 |  sampleBy(self, col, fractions, seed=None)
 |      Returns a stratified sample without replacement based on the
 |      fraction given on each stratum.
 |
 |      :param col: column that defines strata
 |      :param fractions:
 |          sampling fraction for each stratum. If a stratum is not
 |          specified, we treat its fraction as zero.
 |      :param seed: random seed
 |      :return: a new DataFrame that represents the stratified sample
 |
 |      >>> from pyspark.sql.functions import col
 |      >>> dataset = sqlContext.range(0, 100).select((col("id") % 3).a
lias("key"))
 |      >>> sampled = dataset.sampleBy("key", fractions={0: 0.1, 1: 0.
2}, seed=0)
 |      >>> sampled.groupBy("key").count().orderBy("key").show()
 |      +---+-----+
 |      |key|count|
 |      +---+-----+
 |      |  0|    3|
 |      |  1|    8|
 |      +---+-----+
 |
 |      .. versionadded:: 1.5
 |
 |  save(self, path=None, source=None, mode='error', **options)
 |      Saves the contents of the :class:`DataFrame` to a data source.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameWriter.save` i
nstead.
 |
 |      .. versionadded:: 1.3
 |
 |  saveAsParquetFile(self, path)
 |      Saves the contents as a Parquet file, preserving the schema.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameWriter.parquet
` instead.
 |
 |  saveAsTable(self, tableName, source=None, mode='error', **options)
 |      Saves the contents of this :class:`DataFrame` to a data source
 as a table.
 |
 |      .. note:: Deprecated in 1.4, use :func:`DataFrameWriter.saveAsT
able` instead.
 |
 |  select(self, *cols)
 |      Projects a set of expressions and returns a new :class:`DataFra
me`.
 |
 |      :param cols: list of column names (string) or expressions (:cla
ss:`Column`).
 |          If one of the column names is '*', that column is expanded
 to include all columns
```

```
 |          in the current DataFrame.
 |
 |      >>> df.select('*').collect()
 |      [Row(age=2, name=u'Alice'), Row(age=5, name=u'Bob')]
 |      >>> df.select('name', 'age').collect()
 |      [Row(name=u'Alice', age=2), Row(name=u'Bob', age=5)]
 |      >>> df.select(df.name, (df.age + 10).alias('age')).collect()
 |      [Row(name=u'Alice', age=12), Row(name=u'Bob', age=15)]
 |
 |      .. versionadded:: 1.3
 |
 |  selectExpr(self, *expr)
 |      Projects a set of SQL expressions and returns a new :class:`Dat
aFrame`.
 |
 |      This is a variant of :func:`select` that accepts SQL expression
s.
 |
 |      >>> df.selectExpr("age * 2", "abs(age)").collect()
 |      [Row((age * 2)=4, 'abs(age)=2), Row((age * 2)=10, 'abs(age)=5)]
 |
 |      .. versionadded:: 1.3
 |
 |  show(self, n=20, truncate=True)
 |      Prints the first ``n`` rows to the console.
 |
 |      :param n: Number of rows to show.
 |      :param truncate: Whether truncate long strings and align cells
 right.
 |
 |      >>> df
 |      DataFrame[age: int, name: string]
 |      >>> df.show()
 |      +---+-----+
 |      |age| name|
 |      +---+-----+
 |      |  2|Alice|
 |      |  5|  Bob|
 |      +---+-----+
 |
 |      .. versionadded:: 1.3
 |
 |  sort(self, *cols, **kwargs)
 |      Returns a new :class:`DataFrame` sorted by the specified column
(s).
 |
 |      :param cols: list of :class:`Column` or column names to sort b
y.
 |      :param ascending: boolean or list of boolean (default True).
 |          Sort ascending vs. descending. Specify list for multiple so
rt orders.
 |          If a list is specified, length of the list must equal lengt
h of the `cols`.
 |
 |      >>> df.sort(df.age.desc()).collect()
 |      [Row(age=5, name=u'Bob'), Row(age=2, name=u'Alice')]
 |      >>> df.sort("age", ascending=False).collect()
```

```
 |      [Row(age=5, name=u'Bob'), Row(age=2, name=u'Alice')]
 |      >>> df.orderBy(df.age.desc()).collect()
 |      [Row(age=5, name=u'Bob'), Row(age=2, name=u'Alice')]
 |      >>> from pyspark.sql.functions import *
 |      >>> df.sort(asc("age")).collect()
 |      [Row(age=2, name=u'Alice'), Row(age=5, name=u'Bob')]
 |      >>> df.orderBy(desc("age"), "name").collect()
 |      [Row(age=5, name=u'Bob'), Row(age=2, name=u'Alice')]
 |      >>> df.orderBy(["age", "name"], ascending=[0, 1]).collect()
 |      [Row(age=5, name=u'Bob'), Row(age=2, name=u'Alice')]
 |
 |      .. versionadded:: 1.3
 |
 |  subtract(self, other)
 |      Return a new :class:`DataFrame` containing rows in this frame
 |      but not in another frame.
 |
 |      This is equivalent to `EXCEPT` in SQL.
 |
 |      .. versionadded:: 1.3
 |
 |  take(self, num)
 |      Returns the first ``num`` rows as a :class:`list` of :class:`Ro
w`.
 |
 |      >>> df.take(2)
 |      [Row(age=2, name=u'Alice'), Row(age=5, name=u'Bob')]
 |
 |      .. versionadded:: 1.3
 |
 |  toJSON(self, use_unicode=True)
 |      Converts a :class:`DataFrame` into a :class:`RDD` of string.
 |
 |      Each row is turned into a JSON document as one element in the r
eturned RDD.
 |
 |      >>> df.toJSON().first()
 |      u'{"age":2,"name":"Alice"}'
 |
 |      .. versionadded:: 1.3
 |
 |  toPandas(self)
 |      Returns the contents of this :class:`DataFrame` as Pandas ``pan
das.DataFrame``.
 |
 |      This is only available if Pandas is installed and available.
 |
 |      >>> df.toPandas()  # doctest: +SKIP
 |         age   name
 |      0    2  Alice
 |      1    5    Bob
 |
 |      .. versionadded:: 1.3
 |
 |  unionAll(self, other)
 |      Return a new :class:`DataFrame` containing union of rows in thi
s
```

```
 |      frame and another frame.
 |      
 |      This is equivalent to `UNION ALL` in SQL.
 |      
 |      .. versionadded:: 1.3
 |  
 |  unpersist(self, blocking=True)
 |      Marks the :class:`DataFrame` as non-persistent, and remove all
 blocks for it from
 |      memory and disk.
 |      
 |      .. versionadded:: 1.3
 |  
 |  where = filter(self, condition)
 |  
 |  withColumn(self, colName, col)
 |      Returns a new :class:`DataFrame` by adding a column or replacin
g the
 |      existing column that has the same name.
 |      
 |      :param colName: string, name of the new column.
 |      :param col: a :class:`Column` expression for the new column.
 |      
 |      >>> df.withColumn('age2', df.age + 2).collect()
 |      [Row(age=2, name=u'Alice', age2=4), Row(age=5, name=u'Bob', age
2=7)]
 |      
 |      .. versionadded:: 1.3
 |  
 |  withColumnRenamed(self, existing, new)
 |      Returns a new :class:`DataFrame` by renaming an existing colum
n.
 |      
 |      :param existing: string, name of the existing column to rename.
 |      :param col: string, new name of the column.
 |      
 |      >>> df.withColumnRenamed('age', 'age2').collect()
 |      [Row(age2=2, name=u'Alice'), Row(age2=5, name=u'Bob')]
 |      
 |      .. versionadded:: 1.3
 |  
 |  ----------------------------------------------------------------------
---
 |  Data descriptors defined here:
 |  
 |  __dict__
 |      dictionary for instance variables (if defined)
 |  
 |  __weakref__
 |      list of weak references to the object (if defined)
 |  
 |  columns
 |      Returns all column names as a list.
 |      
 |      >>> df.columns
 |      ['age', 'name']
 |      
```

```
 |      .. versionadded:: 1.3
 |
 |  dtypes
 |      Returns all column names and their data types as a list.
 |
 |      >>> df.dtypes
 |      [('age', 'int'), ('name', 'string')]
 |
 |      .. versionadded:: 1.3
 |
 |  na
 |      Returns a :class:`DataFrameNaFunctions` for handling missing va
lues.
 |
 |      .. versionadded:: 1.3.1
 |
 |  rdd
 |      Returns the content as an :class:`pyspark.RDD` of :class:`Row`.
 |
 |      .. versionadded:: 1.3
 |
 |  schema
 |      Returns the schema of this :class:`DataFrame` as a :class:`type
s.StructType`.
 |
 |      >>> df.schema
 |      StructType(List(StructField(age,IntegerType,true),StructField(n
ame,StringType,true)))
 |
 |      .. versionadded:: 1.3
 |
 |  stat
 |      Returns a :class:`DataFrameStatFunctions` for statistic functio
ns.
 |
 |      .. versionadded:: 1.4
 |
 |  write
 |      Interface for saving the content of the :class:`DataFrame` out
 into external storage.
 |
 |      :return: :class:`DataFrameWriter`
 |
 |      .. versionadded:: 1.4
```

How many partitions will the DataFrame be split into?

In [24]:
```
dataDF.rdd.getNumPartitions()
```

Out[24]: 128

A note about DataFrames and queries

When you use DataFrames or Spark SQL, you are building up a *query plan*. Each transformation you apply to a DataFrame adds some information to the query plan. When you finally call an action, which triggers execution of your Spark job, several things happen:

1. Spark's Catalyst optimizer analyzes the query plan (called an *unoptimized logical query plan*) and attempts to optimize it. Optimizations include (but aren't limited to) rearranging and combining `filter()` operations for efficiency, converting `Decimal` operations to more efficient long integer operations, and pushing some operations down into the data source (e.g., a `filter()` operation might be translated to a SQL `WHERE` clause, if the data source is a traditional SQL RDBMS). The result of this optimization phase is an *optimized logical plan*.
2. Once Catalyst has an optimized logical plan, it then constructs multiple *physical* plans from it. Specifically, it implements the query in terms of lower level Spark RDD operations.
3. Catalyst chooses which physical plan to use via *cost optimization*. That is, it determines which physical plan is the most efficient (or least expensive), and uses that one.
4. Finally, once the physical RDD execution plan is established, Spark actually executes the job.

You can examine the query plan using the `explain()` function on a DataFrame. By default, `explain()` only shows you the final physical plan; however, if you pass it an argument of `True`, it will show you all phases.

(If you want to take a deeper dive into how Catalyst optimizes DataFrame queries, this blog post, while a little old, is an excellent overview: [Deep Dive into Spark SQL's Catalyst Optimizer (https://databricks.com/blog/2015/04/13/deep-dive-into-spark-sqls-catalyst-optimizer.html)](https://databricks.com/blog/2015/04/13/deep-dive-into-spark-sqls-catalyst-optimizer.html).)

Let's add a couple transformations to our DataFrame and look at the query plan on the resulting transformed DataFrame. Don't be too concerned if it looks like gibberish. As you gain more experience with Apache Spark, you'll begin to be able to use `explain()` to help you understand more about your DataFrame operations.

In [25]:
```
newDF = dataDF.distinct().select('*')
newDF.explain(True)
```

```
== Parsed Logical Plan ==
'Project [*]
 Aggregate [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], [last_
name#0,first_name#1,ssn#2,occupation#3,age#4L]
  LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapP
artitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.j
ava:-2

== Analyzed Logical Plan ==
last_name: string, first_name: string, ssn: string, occupation: string,
age: bigint
Project [last_name#0,first_name#1,ssn#2,occupation#3,age#4L]
 Aggregate [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], [last_
name#0,first_name#1,ssn#2,occupation#3,age#4L]
  LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapP
artitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.j
ava:-2

== Optimized Logical Plan ==
Aggregate [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], [last_n
ame#0,first_name#1,ssn#2,occupation#3,age#4L]
 LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapPa
rtitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.ja
va:-2

== Physical Plan ==
TungstenAggregate(key=[last_name#0,first_name#1,ssn#2,occupation#3,age#
4L], functions=[], output=[last_name#0,first_name#1,ssn#2,occupation#3,
age#4L])
 TungstenExchange hashpartitioning(last_name#0,first_name#1,ssn#2,occup
ation#3,age#4L)
  TungstenAggregate(key=[last_name#0,first_name#1,ssn#2,occupation#3,ag
e#4L], functions=[], output=[last_name#0,first_name#1,ssn#2,occupation#
3,age#4L])
   Scan PhysicalRDD[last_name#0,first_name#1,ssn#2,occupation#3,age#4L]

Code Generation: true
```

(3c): Subtract one from each value using *select*

So far, we've created a distributed DataFrame that is split into many partitions, where each partition is stored on a single machine in our cluster. Let's look at what happens when we do a basic operation on the dataset. Many useful data analysis operations can be specified as "do something to each item in the dataset". These data-parallel operations are convenient because each item in the dataset can be processed individually: the operation on one entry doesn't effect the operations on any of the other entries. Therefore, Spark can parallelize the operation.

One of the most common DataFrame operations is `select()`, and it works more or less like a SQL `SELECT` statement: You can select specific columns from the DataFrame, and you can even use `select()` to create *new* columns with values that are derived from existing column values. We can use `select()` to create a new column that decrements the value of the existing `age` column.

`select()` is a *transformation*. It returns a new DataFrame that captures both the previous DataFrame and the operation to add to the query (`select`, in this case). But it does *not* actually execute anything on the cluster. When transforming DataFrames, we are building up a *query plan*. That query plan will be optimized, implemented (in terms of RDDs), and executed by Spark *only* when we call an action.

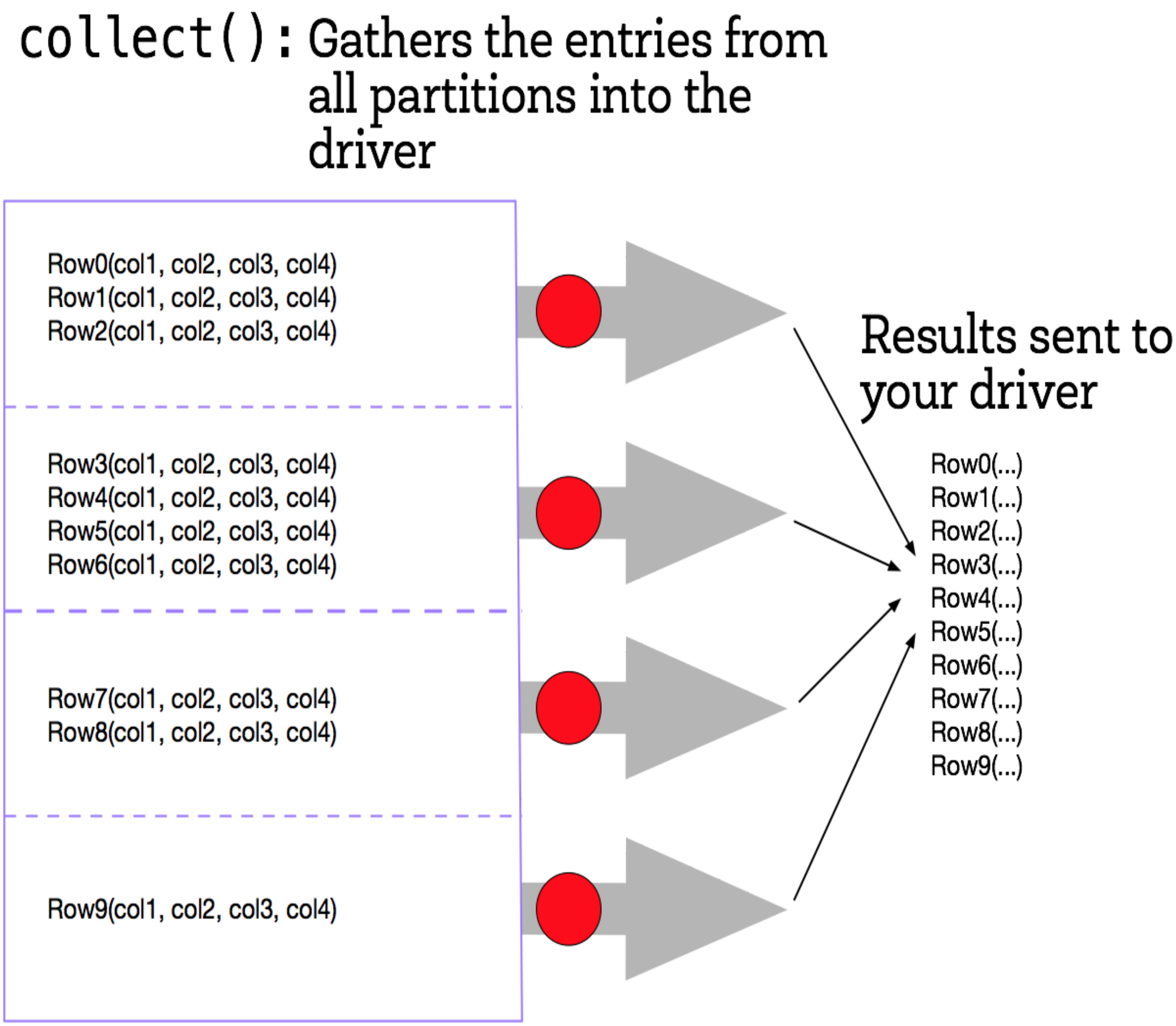
In [26]:
```
# Transform dataDF through a select transformation and rename the newly
created '(age -1)' column to 'age'
# Because select is a transformation and Spark uses lazy evaluation, no
jobs, stages,
# or tasks will be launched when we run this code.
subDF = dataDF.select('last_name', 'first_name', 'ssn', 'occupation',
(dataDF.age - 1).alias('age'))
```

Let's take a look at the query plan.

In [27]:
```
subDF.explain(True)
```

```
== Parsed Logical Plan ==
'Project [unresolvedalias('last_name),unresolvedalias('first_name),unre
solvedalias('ssn),unresolvedalias('occupation),(age#4L - 1) AS age#5]
 LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapPa
rtitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.ja
va:-2

== Analyzed Logical Plan ==
last_name: string, first_name: string, ssn: string, occupation: string,
age: bigint
Project [last_name#0,first_name#1,ssn#2,occupation#3,(age#4L - cast(1 a
s bigint)) AS age#5L]
 LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapPa
rtitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.ja
va:-2

== Optimized Logical Plan ==
Project [last_name#0,first_name#1,ssn#2,occupation#3,(age#4L - 1) AS ag
e#5L]
 LogicalRDD [last_name#0,first_name#1,ssn#2,occupation#3,age#4L], MapPa
rtitionsRDD[5] at applySchemaToPythonRDD at NativeMethodAccessorImpl.ja
va:-2

== Physical Plan ==
TungstenProject [last_name#0,first_name#1,ssn#2,occupation#3,(age#4L -
 1) AS age#5L]
 Scan PhysicalRDD[last_name#0,first_name#1,ssn#2,occupation#3,age#4L]

Code Generation: true
```

(3d) Use *collect* to view results



To see a list of elements decremented by one, we need to create a new list on the driver from the the data distributed in the executor nodes. To do this we can call the `collect()` method on our DataFrame. `collect()` is often used after transformations to ensure that we are only returning a *small* amount of data to the driver. This is done because the data returned to the driver must fit into the driver's available memory. If not, the driver will crash.

The `collect()` method is the first action operation that we have encountered. Action operations cause Spark to perform the (lazy) transformation operations that are required to compute the values returned by the action. In our example, this means that tasks will now be launched to perform the `createDataFrame`, `select`, and `collect` operations.

In the diagram, the dataset is broken into four partitions, so four `collect()` tasks are launched. Each task collects the entries in its partition and sends the result to the driver, which creates a list of the values, as shown in the figure below.

Now let's run `collect()` on `subDF`.

In [28]:

```
# Let's collect the data
results = subDF.collect()
print results
```

```
[Row(last_name=u'Brown', first_name=u'Jason', ssn=u'160-37-9051', occup
ation=u'Agricultural engineer', age=38), Row(last_name=u'Morrison', fir
st_name=u'Earl', ssn=u'361-94-4342', occupation=u'Teacher, primary scho
ol', age=25), Row(last_name=u'Young', first_name=u'Christian', ssn=u'76
9-27-5887', occupation=u'Scientific laboratory technician', age=20), Ro
w(last_name=u'George', first_name=u'Daniel', ssn=u'175-24-7915', occupa
tion=u'Geophysicist/field seismologist', age=41), Row(last_name=u'Lee',
first_name=u'Dawn', ssn=u'310-69-7326', occupation=u'Forensic psycholog
ist', age=25), Row(last_name=u'Hamilton', first_name=u'Matthew', ssn=
u'099-90-9730', occupation=u'Best boy', age=42), Row(last_name=u'Amand
a', first_name=u'Miss', ssn=u'476-06-5497', occupation=u'English as a f
oreign language teacher', age=42), Row(last_name=u'Fernandez', first_na
me=u'Thomas', ssn=u'722-09-8354', occupation=u'Psychologist, prison and
probation services', age=5), Row(last_name=u'Kirk', first_name=u'Jennif
er', ssn=u'715-56-1708', occupation=u'Sales executive', age=4), Row(las
t_name=u'Young', first_name=u'Todd', ssn=u'123-48-8354', occupation=u'E
ngineer, broadcasting (operations)', age=16), Row(last_name=u'Cook', fi
rst_name=u'Amy', ssn=u'293-22-0265', occupation=u'Scientist, product/pr
ocess development', age=27), Row(last_name=u'Martinez', first_name=u'Ma
rk', ssn=u'041-23-3263', occupation=u'Building control surveyor', age=2
3), Row(last_name=u'Walker', first_name=u'Aaron', ssn=u'725-61-1132', o
ccupation=u'Artist', age=42), Row(last_name=u'Dennis', first_name=u'Edw
ard', ssn=u'268-79-4330', occupation=u'Chiropractor', age=14), Row(last
_name=u'Hebert', first_name=u'Meredith', ssn=u'077-96-8349', occupation
=u'Surveyor, minerals', age=15), Row(last_name=u'Hess', first_name=u'Ph
yllis', ssn=u'061-88-1648', occupation=u'Production assistant, televisi
on', age=32), Row(last_name=u'Reed', first_name=u'Kayla', ssn=u'582-28-
0099', occupation=u'Manufacturing systems engineer', age=11), Row(last_
name=u'Miller', first_name=u'Dwayne', ssn=u'386-07-6013', occupation=
u'Social research officer, government', age=5), Row(last_name=u'Weeks',
first_name=u'Sean', ssn=u'363-94-7993', occupation=u'Administrator, edu
cation', age=41), Row(last_name=u'Williams', first_name=u'Cassie', ssn=
u'386-39-5490', occupation=u'Horticulturist, commercial', age=10), Row
(last_name=u'Rose', first_name=u'Daniel', ssn=u'737-44-0894', occupatio
n=u'Economist', age=7), Row(last_name=u'Dawson', first_name=u'Taylor',
 ssn=u'790-03-8999', occupation=u'Psychotherapist', age=25), Row(last_n
ame=u'Fisher', first_name=u'Steven', ssn=u'789-24-6522', occupation=u'T
herapist, drama', age=33), Row(last_name=u'Johnson', first_name=u'Steph
anie', ssn=u'702-94-4924', occupation=u'Press sub', age=39), Row(last_n
ame=u'Farmer', first_name=u'Donna', ssn=u'117-61-4564', occupation=u'Mu
sic tutor', age=14), Row(last_name=u'Davis', first_name=u'Tonya', ssn=
u'695-93-4517', occupation=u'Conference centre manager', age=34), Row(l
ast_name=u'Robert', first_name=u'Dr.', ssn=u'121-52-8368', occupation=
u'Buyer, industrial', age=16), Row(last_name=u'Shaw', first_name=u'Davi
d', ssn=u'672-57-1650', occupation=u'Programmer, multimedia', age=23),
 Row(last_name=u'Paul', first_name=u'Victoria', ssn=u'370-63-8920', occ
upation=u'Exercise physiologist', age=21), Row(last_name=u'Hays', first
_name=u'Jeff', ssn=u'661-20-4642', occupation=u'Nature conservation off
icer', age=21), Row(last_name=u'Nguyen', first_name=u'Erica', ssn=u'251
-56-0999', occupation=u'Agricultural engineer', age=41), Row(last_name=
u'Skinner', first_name=u'Christopher', ssn=u'621-23-3079', occupation=
u'Ophthalmologist', age=45), Row(last_name=u'Jackson', first_name=u'Nat
han', ssn=u'682-09-3573', occupation=u'Administrator, local governmen
t', age=34), Row(last_name=u'Thompson', first_name=u'Terri', ssn=u'607-
24-8694', occupation=u'Furniture designer', age=32), Row(last_name=u'Fi
gueroa', first_name=u'Ashley', ssn=u'439-10-9336', occupation=u'Geograp
hical information systems officer', age=20), Row(last_name=u'Lewis', fi
```

rst_name=u'David', ssn=u'695-60-0033', occupation=u'Financial risk anal
yst', age=9), Row(last_name=u'Rodgers', first_name=u'Beverly', ssn=u'81
9-57-7194', occupation=u'Regulatory affairs officer', age=0), Row(last_
name=u'Randall', first_name=u'Dennis', ssn=u'077-40-8314', occupation=
u'Lexicographer', age=25), Row(last_name=u'Quinn', first_name=u'Teres
a', ssn=u'823-43-5986', occupation=u'Human resources officer', age=34),
Row(last_name=u'Ford', first_name=u'Joshua', ssn=u'623-03-3910', occupa
tion=u'Editorial assistant', age=28), Row(last_name=u'Watts', first_nam
e=u'Joseph', ssn=u'225-30-2177', occupation=u'Exercise physiologist', a
ge=25), Row(last_name=u'Martin', first_name=u'Willie', ssn=u'174-81-653
5', occupation=u'Engineer, manufacturing systems', age=44), Row(last_na
me=u'Davis', first_name=u'Joseph', ssn=u'874-11-0910', occupation=u'Car
eers adviser', age=16), Row(last_name=u'Rodriguez', first_name=u'Ross',
ssn=u'275-26-3828', occupation=u'Production manager', age=38), Row(last
_name=u'Rubio', first_name=u'Christopher', ssn=u'750-41-6999', occupati
on=u'Surveyor, quantity', age=10), Row(last_name=u'Johnson', first_name
=u'Amanda', ssn=u'494-58-8536', occupation=u'Horticulturist, amenity',
age=3), Row(last_name=u'Haas', first_name=u'Emily', ssn=u'518-20-458
7', occupation=u'Retail buyer', age=39), Row(last_name=u'Young', first_
name=u'Eddie', ssn=u'093-34-0390', occupation=u'Dentist', age=21), Row
(last_name=u'Gordon', first_name=u'Shawn', ssn=u'422-63-3984', occupati
on=u'Agricultural engineer', age=21), Row(last_name=u'Wilson', first_na
me=u'Tiffany', ssn=u'560-10-3377', occupation=u'Dramatherapist', age=4
6), Row(last_name=u'Jenkins', first_name=u'Robert', ssn=u'828-24-7280',
occupation=u'Designer, furniture', age=27), Row(last_name=u'Webb', firs
t_name=u'Megan', ssn=u'664-41-7482', occupation=u'Database administrato
r', age=37), Row(last_name=u'Lewis', first_name=u'Jaime', ssn=u'340-54-
3457', occupation=u'Legal executive', age=36), Row(last_name=u'Ford', f
irst_name=u'Kaitlyn', ssn=u'017-37-9292', occupation=u'Designer, exhibi
tion/display', age=27), Row(last_name=u'Hall', first_name=u'Monique', s
sn=u'518-71-6777', occupation=u'Restaurant manager, fast food', age=3
7), Row(last_name=u'Mendoza', first_name=u'Michael', ssn=u'227-93-168
6', occupation=u'Surveyor, building', age=30), Row(last_name=u'Johnso
n', first_name=u'Nicholas', ssn=u'157-13-2050', occupation=u'Journalis
t, broadcasting', age=0), Row(last_name=u'Ball', first_name=u'Jacquelin
e', ssn=u'604-21-5781', occupation=u'Commercial horticulturist', age=1
8), Row(last_name=u'Walker', first_name=u'Joseph', ssn=u'079-99-5775',
occupation=u'Retail manager', age=28), Row(last_name=u'Hernandez', fir
st_name=u'Richard', ssn=u'639-98-4417', occupation=u'Event organiser',
age=27), Row(last_name=u'Munoz', first_name=u'Kayla', ssn=u'784-13-101
4', occupation=u'Ergonomist', age=32), Row(last_name=u'Fry', first_name
=u'Jeffrey', ssn=u'337-29-9383', occupation=u'Careers adviser', age=3
6), Row(last_name=u'Preston', first_name=u'Tyler', ssn=u'885-52-6902',
occupation=u'Production engineer', age=35), Row(last_name=u'Murphy', f
irst_name=u'Lisa', ssn=u'475-39-0779', occupation=u'Ranger/warden', age
=45), Row(last_name=u'Johnson', first_name=u'Candice', ssn=u'639-80-684
1', occupation=u'Engineer, automotive', age=5), Row(last_name=u'Neal',
first_name=u'Amanda', ssn=u'685-37-7370', occupation=u'Tourist informa
tion centre manager', age=38), Row(last_name=u'Kaufman', first_name=u'M
ichael', ssn=u'258-69-6049', occupation=u'Chemical engineer', age=23),
Row(last_name=u'Smith', first_name=u'Ashlee', ssn=u'387-51-5783', occu
pation=u'Teacher, primary school', age=17), Row(last_name=u'Arias', fir
st_name=u'Mark', ssn=u'501-67-7559', occupation=u'Marine scientist', ag
e=3), Row(last_name=u'Webster', first_name=u'Jennifer', ssn=u'156-76-67
78', occupation=u'Ergonomist', age=33), Row(last_name=u'Rose', first_na
me=u'Henry', ssn=u'314-03-4053', occupation=u'Broadcast journalist', ag
e=25), Row(last_name=u'Barry', first_name=u'Jeremy', ssn=u'004-41-324

3', occupation=u'Accommodation manager', age=15), Row(last_name=u'Jone
s', first_name=u'Jeremy', ssn=u'648-72-8898', occupation=u'Production d
esigner, theatre/television/film', age=8), Row(last_name=u'David', firs
t_name=u'Tracy', ssn=u'384-34-4827', occupation=u'Lecturer, higher educ
ation', age=37), Row(last_name=u'Ibarra', first_name=u'Anthony', ssn=
u'843-46-1528', occupation=u'Designer, interior/spatial', age=15), Row
(last_name=u'Larson', first_name=u'Joseph', ssn=u'076-36-8031', occupat
ion=u'Engineer, biomedical', age=26), Row(last_name=u'Turner', first_na
me=u'Ariana', ssn=u'176-16-1892', occupation=u'Leisure centre manager',
age=38), Row(last_name=u'Delgado', first_name=u'Sonia', ssn=u'283-58-49
51', occupation=u'Special educational needs teacher', age=6), Row(last_
name=u'Bishop', first_name=u'Jonathan', ssn=u'714-87-5335', occupation=
u'Medical technical officer', age=10), Row(last_name=u'Pineda', first_n
ame=u'Connor', ssn=u'041-63-6459', occupation=u'Designer, jewellery', a
ge=10), Row(last_name=u'Craig', first_name=u'Jacob', ssn=u'656-19-987
5', occupation=u'Control and instrumentation engineer', age=28), Row(la
st_name=u'Robinson', first_name=u'Adam', ssn=u'279-56-1630', occupation
=u'Fish farm manager', age=1), Row(last_name=u'Holloway', first_name=
u'Julia', ssn=u'115-33-9587', occupation=u'Animal technologist', age=
3), Row(last_name=u'Fleming', first_name=u'Jodi', ssn=u'833-40-8181', o
ccupation=u'Agricultural consultant', age=0), Row(last_name=u'Powell',
first_name=u'Kimberly', ssn=u'102-37-5719', occupation=u'Counsellor',
age=6), Row(last_name=u'Gamble', first_name=u'Julie', ssn=u'374-87-437
2', occupation=u'Electrical engineer', age=17), Row(last_name=u'Brenna
n', first_name=u'Christopher', ssn=u'831-32-0206', occupation=u'Secreta
ry/administrator', age=43), Row(last_name=u'Green', first_name=u'Kare
n', ssn=u'531-17-1551', occupation=u'Optician, dispensing', age=22), Ro
w(last_name=u'Cook', first_name=u'Jessica', ssn=u'811-23-6545', occupat
ion=u'Toxicologist', age=23), Row(last_name=u'Walker', first_name=u'Rho
nda', ssn=u'648-03-3938', occupation=u'Scientist, marine', age=23), Row
(last_name=u'Perez', first_name=u'Elizabeth', ssn=u'571-87-0459', occup
ation=u'Sports therapist', age=42), Row(last_name=u'Valencia', first_na
me=u'Alicia', ssn=u'274-89-5763', occupation=u'Leisure centre manager',
age=5), Row(last_name=u'Williams', first_name=u'Brandon', ssn=u'590-16-
3719', occupation=u'Editor, film/video', age=16), Row(last_name=u'Hun
t', first_name=u'Michael', ssn=u'745-13-3849', occupation=u'Sales profe
ssional, IT', age=24), Row(last_name=u'Sandoval', first_name=u'Catherin
e', ssn=u'295-44-6367', occupation=u'Lecturer, higher education', age=4
6), Row(last_name=u'Stephenson', first_name=u'Ronald', ssn=u'659-93-662
2', occupation=u'Armed forces training and education officer', age=32),
Row(last_name=u'Barton', first_name=u'Ryan', ssn=u'728-17-7313', occupa
tion=u'Art gallery manager', age=25), Row(last_name=u'Smith', first_nam
e=u'Rachel', ssn=u'708-44-5175', occupation=u'Surveyor, commercial/resi
dential', age=38), Row(last_name=u'Griffith', first_name=u'Valerie', ss
n=u'182-03-1388', occupation=u'Field trials officer', age=6), Row(last_
name=u'Jordan', first_name=u'Ashley', ssn=u'399-94-7975', occupation=
u'Educational psychologist', age=31), Row(last_name=u'Price', first_nam
e=u'Ronald', ssn=u'694-58-1212', occupation=u'Sales professional, IT',
age=24), Row(last_name=u'Esparza', first_name=u'John', ssn=u'607-25-07
33', occupation=u'Therapeutic radiographer', age=29), Row(last_name=u'P
eck', first_name=u'Douglas', ssn=u'133-19-9920', occupation=u'Senior ta
x professional/tax inspector', age=2), Row(last_name=u'Black', first_na
me=u'Kimberly', ssn=u'677-92-2450', occupation=u'Chemical engineer', ag
e=13), Row(last_name=u'Charles', first_name=u'Mr.', ssn=u'323-86-6963',
occupation=u'Engineer, energy', age=18), Row(last_name=u'Moore', first_
name=u'David', ssn=u'133-48-6552', occupation=u'Surveyor, land/geomatic
s', age=36), Row(last_name=u'Ramirez', first_name=u'Christine', ssn=u'0

```
87-21-9165', occupation=u'Secretary/administrator', age=24), Row(last_n
ame=u'Morrow', first_name=u'Alex', ssn=u'533-54-4033', occupation=u'Tox
icologist', age=3), Row(last_name=u'Johnson', first_name=u'Christian',
 ssn=u'860-63-4089', occupation=u'Medical technical officer', age=16),
 Row(last_name=u'Blair', first_name=u'Steven', ssn=u'675-95-9698', occu
pation=u'Personal assistant', age=33), Row(last_name=u'Jones', first_na
me=u'Natalie', ssn=u'429-52-5642', occupation=u'Restaurant manager, fas
t food', age=21), Row(last_name=u'Cummings', first_name=u'Jeanette', ss
n=u'149-57-3736', occupation=u'Broadcast journalist', age=15), Row(last
_name=u'Payne', first_name=u'Jacob', ssn=u'246-27-8489', occupation=u'F
reight forwarder', age=35), Row(last_name=u'Williams', first_name=u'Pat
rick', ssn=u'812-88-9795', occupation=u'Programmer, systems', age=46),
 Row(last_name=u'Meyer', first_name=u'Cristian', ssn=u'602-45-1203', oc
cupation=u'Geographical information systems officer', age=7), Row(last_
name=u'Murphy', first_name=u'Michael', ssn=u'663-26-5493', occupation=
u'Surveyor, planning and development', age=7), Row(last_name=u'Carlso
n', first_name=u'Jerry', ssn=u'823-88-1208', occupation=u'Art therapis
t', age=20), Row(last_name=u'Schultz', first_name=u'Victoria', ssn=u'15
9-58-7310', occupation=u'Research officer, political party', age=44), R
ow(last_name=u'Jackson', first_name=u'Cynthia', ssn=u'699-10-4819', occ
upation=u'Furniture conservator/restorer', age=42), Row(last_name=u'Mil
ler', first_name=u'Jacob', ssn=u'691-69-2880', occupation=u'Civil engin
eer, consulting', age=29), Row(last_name=u'Wang', first_name=u'Edward',
ssn=u'761-51-6731', occupation=u'Training and development officer', age
=14), Row(last_name=u'Smith', first_name=u'Christopher', ssn=u'409-86-2
882', occupation=u'Equality and diversity officer', age=5), Row(last_na
me=u'Thomas', first_name=u'Breanna', ssn=u'700-90-3741', occupation=u'C
ommunity arts worker', age=44), Row(last_name=u'Roberts', first_name=
u'Luis', ssn=u'562-33-8154', occupation=u'Psychologist, occupational',
 age=11), Row(last_name=u'Davis', first_name=u'Briana', ssn=u'687-12-38
00', occupation=u'Research scientist (medical)', age=24), Row(last_name
=u'Arias', first_name=u'Jessica', ssn=u'320-58-5487', occupation=u'Occu
pational psychologist', age=45), Row(last_name=u'Johnson', first_name=
u'Hector', ssn=u'753-83-1027', occupation=u'Systems analyst', age=16),
 Row(last_name=u'Gallegos', first_name=u'Michael', ssn=u'203-15-1520',
 occupation=u'Water engineer', age=0), Row(last_name=u'George', first_n
ame=u'Kathryn', ssn=u'068-01-3503', occupation=u'Development worker, in
ternational aid', age=34), Row(last_name=u'Moore', first_name=u'Scott',
ssn=u'514-35-9041', occupation=u'Public relations officer', age=24), Ro
w(last_name=u'Frederick', first_name=u'Anthony', ssn=u'607-75-8076', oc
cupation=u'Ceramics designer', age=27), Row(last_name=u'Evans', first_n
ame=u'Raven', ssn=u'829-56-8556', occupation=u'Heritage manager', age=
9), Row(last_name=u'Washington', first_name=u'Gregory', ssn=u'325-60-01
18', occupation=u'Advertising account planner', age=46), Row(last_name=
u'Logan', first_name=u'Elizabeth', ssn=u'510-11-4273', occupation=u'Res
earch officer, government', age=31), Row(last_name=u'King', first_name=
u'Jacqueline', ssn=u'506-33-4081', occupation=u'Sales executive', age=
1), Row(last_name=u'Cook', first_name=u'Ryan', ssn=u'460-83-5028', occu
pation=u'Brewing technologist', age=2), Row(last_name=u'Marsh', first_n
ame=u'Destiny', ssn=u'560-35-4759', occupation=u'Hydrologist', age=38),
Row(last_name=u'Luna', first_name=u'Peter', ssn=u'186-20-8593', occupat
ion=u'Scientist, clinical (histocompatibility and immunogenetics)', age
=20), Row(last_name=u'Rios', first_name=u'Jennifer', ssn=u'212-15-419
9', occupation=u'Claims inspector/assessor', age=24), Row(last_name=u'J
ohnson', first_name=u'Holly', ssn=u'093-03-1883', occupation=u'Insuranc
e account manager', age=1), Row(last_name=u'Barron', first_name=u'Court
ney', ssn=u'519-80-0824', occupation=u'Engineer, drilling', age=10), Ro
```

w(last_name=u'Schneider', first_name=u'Jessica', ssn=u'794-09-9597', occupation=u'Environmental manager', age=25), Row(last_name=u'Wright', first_name=u'Roger', ssn=u'287-49-5958', occupation=u'Electronics engineer', age=12), Row(last_name=u'Watkins', first_name=u'Isaac', ssn=u'698-36-7612', occupation=u'Quarry manager', age=15), Row(last_name=u'Gonzalez', first_name=u'Stephanie', ssn=u'880-98-8223', occupation=u'Teacher, secondary school', age=38), Row(last_name=u'Hughes', first_name=u'Amy', ssn=u'255-17-3490', occupation=u'Chemical engineer', age=10), Row(last_name=u'Marshall', first_name=u'Geoffrey', ssn=u'826-82-5735', occupation=u'Development worker, community', age=1), Row(last_name=u'Munoz', first_name=u'Denise', ssn=u'617-10-3267', occupation=u'Passenger transport manager', age=31), Row(last_name=u'Rodriguez', first_name=u'Laura', ssn=u'380-71-1131', occupation=u'Financial risk analyst', age=6), Row(last_name=u'Nelson', first_name=u'William', ssn=u'618-59-7628', occupation=u'Scientific laboratory technician', age=38), Row(last_name=u'Peterson', first_name=u'Katherine', ssn=u'119-45-7899', occupation=u'Diagnostic radiographer', age=27), Row(last_name=u'Williams', first_name=u'Nicole', ssn=u'574-72-0358', occupation=u'Advice worker', age=4), Row(last_name=u'Hubbard', first_name=u'Travis', ssn=u'202-29-0251', occupation=u'Plant breeder/geneticist', age=12), Row(last_name=u'Williams', first_name=u'Jerry', ssn=u'062-08-0637', occupation=u'Environmental manager', age=31), Row(last_name=u'Hood', first_name=u'Haley', ssn=u'754-44-5517', occupation=u'Psychiatrist', age=18), Row(last_name=u'Dorsey', first_name=u'Brandon', ssn=u'862-30-8837', occupation=u'Engineer, broadcasting (operations)', age=21), Row(last_name=u'Wallace', first_name=u'Kenneth', ssn=u'067-92-3736', occupation=u'Programmer, systems', age=22), Row(last_name=u'Lynch', first_name=u'Zachary', ssn=u'622-55-9599', occupation=u'Editorial assistant', age=38), Row(last_name=u'Rodriguez', first_name=u'Donald', ssn=u'873-75-1255', occupation=u'Retail banker', age=39), Row(last_name=u'Newman', first_name=u'Rebecca', ssn=u'050-93-5616', occupation=u'Barrister', age=14), Row(last_name=u'Williams', first_name=u'Blake', ssn=u'248-73-1904', occupation=u'Scientist, physiological', age=32), Row(last_name=u'Martin', first_name=u'Aaron', ssn=u'059-08-0977', occupation=u'Secondary school teacher', age=17), Row(last_name=u'Johnson', first_name=u'Mikayla', ssn=u'746-47-0319', occupation=u'Wellsite geologist', age=36), Row(last_name=u'Bernard', first_name=u'Dawn', ssn=u'307-36-3457', occupation=u'Engineer, electronics', age=21), Row(last_name=u'Gomez', first_name=u'Erin', ssn=u'045-58-9521', occupation=u'Personnel officer', age=16), Row(last_name=u'Mack', first_name=u'Bradley', ssn=u'662-80-6804', occupation=u'Mechanical engineer', age=4), Row(last_name=u'Howard', first_name=u'Nicholas', ssn=u'188-57-7205', occupation=u'Astronomer', age=7), Row(last_name=u'Ellis', first_name=u'George', ssn=u'193-08-3697', occupation=u'Surveyor, planning and development', age=3), Row(last_name=u'Thornton', first_name=u'Heather', ssn=u'265-74-7056', occupation=u'Sound technician, broadcasting/film/video', age=32), Row(last_name=u'Adams', first_name=u'Bobby', ssn=u'869-88-6439', occupation=u'Scientist, research (medical)', age=43), Row(last_name=u'Young', first_name=u'Kristin', ssn=u'033-41-6170', occupation=u'Legal secretary', age=13), Row(last_name=u'Cross', first_name=u'David', ssn=u'856-53-4309', occupation=u'Engineer, agricultural', age=21), Row(last_name=u'Williams', first_name=u'Daniel', ssn=u'318-62-6385', occupation=u'Commercial/residential surveyor', age=38), Row(last_name=u'Quinn', first_name=u'Samuel', ssn=u'143-25-6466', occupation=u'Musician', age=36), Row(last_name=u'Dawson', first_name=u'Kim', ssn=u'096-29-1552', occupation=u'Horticultural therapist', age=39), Row(last_name=u'Horton', first_name=u'Dominic', ssn=u'057-78-9985', occupation=u'Furniture designer', age=17), Row(last_name=u'Solis', first_name=u'Pamela', ssn=u'534-5

1-7738', occupation=u'Travel agency manager', age=19), Row(last_name=
u'Fitzgerald', first_name=u'Benjamin', ssn=u'660-36-5696', occupation=
u'Retail merchandiser', age=45), Row(last_name=u'Lozano', first_name=
u'Steven', ssn=u'234-34-4015', occupation=u'Administrator, sports', age
=34), Row(last_name=u'Bender', first_name=u'George', ssn=u'784-01-371
5', occupation=u'Aid worker', age=17), Row(last_name=u'Ware', first_nam
e=u'Jacqueline', ssn=u'023-58-4110', occupation=u'Exercise physiologis
t', age=28), Row(last_name=u'Collins', first_name=u'Mike', ssn=u'449-82
-1104', occupation=u'Investment banker, operational', age=11), Row(last
_name=u'Martinez', first_name=u'Jeffrey', ssn=u'858-74-5198', occupatio
n=u'Biomedical engineer', age=22), Row(last_name=u'Page', first_name=
u'Keith', ssn=u'654-17-6677', occupation=u'Armed forces operational off
icer', age=25), Row(last_name=u'Smith', first_name=u'Glenn', ssn=u'138-
46-4267', occupation=u'Equities trader', age=3), Row(last_name=u'Rigg
s', first_name=u'Brianna', ssn=u'320-57-4634', occupation=u'Local gover
nment officer', age=30), Row(last_name=u'Martinez', first_name=u'Hanna
h', ssn=u'597-28-2969', occupation=u'Clinical biochemist', age=40), Row
(last_name=u'Smith', first_name=u'Kylie', ssn=u'788-44-1498', occupatio
n=u'Immigration officer', age=33), Row(last_name=u'Ramos', first_name=
u'Matthew', ssn=u'467-37-5167', occupation=u'Air cabin crew', age=44),
 Row(last_name=u'Krueger', first_name=u'Allison', ssn=u'290-22-8699', o
ccupation=u'Programmer, applications', age=19), Row(last_name=u'Avery',
first_name=u'Julia', ssn=u'038-58-5246', occupation=u'Merchant navy off
icer', age=13), Row(last_name=u'Summers', first_name=u'Regina', ssn=u'1
89-39-0401', occupation=u'Training and development officer', age=35), R
ow(last_name=u'Murphy', first_name=u'Jade', ssn=u'654-72-5737', occupat
ion=u'Hotel manager', age=29), Row(last_name=u'Hendricks', first_name=
u'Brianna', ssn=u'861-89-0954', occupation=u'IT trainer', age=11), Row
(last_name=u'Shaw', first_name=u'Jill', ssn=u'131-34-6650', occupation=
u'Television/film/video producer', age=23), Row(last_name=u'Johnson', f
irst_name=u'Patricia', ssn=u'519-66-7678', occupation=u'Claims inspecto
r/assessor', age=24), Row(last_name=u'Galloway', first_name=u'Leah', ss
n=u'746-21-8123', occupation=u'Broadcast engineer', age=45), Row(last_n
ame=u'Walker', first_name=u'Jessica', ssn=u'888-68-3226', occupation=
u'Development worker, international aid', age=28), Row(last_name=u'Gibb
s', first_name=u'Melissa', ssn=u'251-98-6139', occupation=u'Private mus
ic teacher', age=13), Row(last_name=u'Campbell', first_name=u'Rebecca',
ssn=u'182-30-8694', occupation=u'Chartered legal executive (England and
Wales)', age=15), Row(last_name=u'Olson', first_name=u'Nathan', ssn=u'4
71-24-1173', occupation=u'Science writer', age=11), Row(last_name=u'Co
x', first_name=u'Christopher', ssn=u'844-14-5835', occupation=u'Communi
ty education officer', age=19), Row(last_name=u'Ellis', first_name=u'St
even', ssn=u'058-86-5174', occupation=u'Race relations officer', age=
5), Row(last_name=u'Brown', first_name=u'Carly', ssn=u'656-67-2873', oc
cupation=u'Architectural technologist', age=29), Row(last_name=u'Jimene
z', first_name=u'Elizabeth', ssn=u'216-68-7534', occupation=u'Garment/t
extile technologist', age=39), Row(last_name=u'Moore', first_name=u'Mir
anda', ssn=u'822-05-0358', occupation=u'Animal technologist', age=27),
 Row(last_name=u'Bryant', first_name=u'Danielle', ssn=u'686-64-7983', o
ccupation=u'Psychologist, counselling', age=24), Row(last_name=u'Hurs
t', first_name=u'Julie', ssn=u'038-15-7878', occupation=u'Dancer', age=
26), Row(last_name=u'Rhodes', first_name=u'Brittany', ssn=u'335-23-780
3', occupation=u'Insurance underwriter', age=8), Row(last_name=u'Nova
k', first_name=u'Michelle', ssn=u'187-56-8840', occupation=u'Music tuto
r', age=18), Row(last_name=u'Taylor', first_name=u'Frank', ssn=u'071-40
-7625', occupation=u'Logistics and distribution manager', age=37), Row
(last_name=u'Cline', first_name=u'Carolyn', ssn=u'335-29-5530', occupat

```
ion=u'Learning disability nurse', age=17), Row(last_name=u'Watson', fir
st_name=u'Peter', ssn=u'873-15-0152', occupation=u'Minerals surveyor',
 age=14), Row(last_name=u'Andrews', first_name=u'Gregory', ssn=u'424-62
-6913', occupation=u'Youth worker', age=16), Row(last_name=u'Waters', f
irst_name=u'Nicholas', ssn=u'661-38-8878', occupation=u'Biomedical scie
ntist', age=36), Row(last_name=u'Flores', first_name=u'Carla', ssn=u'84
2-51-1940', occupation=u'Food technologist', age=3), Row(last_name=u'Ja
mie', first_name=u'Mrs.', ssn=u'548-90-1314', occupation=u'Financial ad
viser', age=11), Row(last_name=u'Ware', first_name=u'Melissa', ssn=u'39
6-46-1866', occupation=u'Photographer', age=11), Row(last_name=u'Spence
r', first_name=u'Christopher', ssn=u'287-13-2685', occupation=u'Plant b
reeder/geneticist', age=36), Row(last_name=u'Torres', first_name=u'Keit
h', ssn=u'876-82-0195', occupation=u'Trade mark attorney', age=25), Row
(last_name=u'Lewis', first_name=u'Jacqueline', ssn=u'041-54-8809', occu
pation=u'Production assistant, television', age=33), Row(last_name=u'Dr
ake', first_name=u'Andrew', ssn=u'179-91-6586', occupation=u'Merchant n
avy officer', age=2), Row(last_name=u'Rodgers', first_name=u'Kelly', ss
n=u'432-48-2206', occupation=u'Magazine features editor', age=20), Row
(last_name=u'Dunlap', first_name=u'Rebecca', ssn=u'300-88-9795', occupa
tion=u'Geologist, engineering', age=42), Row(last_name=u'Thomas', first
_name=u'Cindy', ssn=u'305-83-7870', occupation=u'Programmer, multimedi
a', age=45), Row(last_name=u'King', first_name=u'Brandon', ssn=u'694-56
-8821', occupation=u'Quarry manager', age=21), Row(last_name=u'Lawrenc
e', first_name=u'Brett', ssn=u'752-13-1070', occupation=u'Air traffic c
ontroller', age=22), Row(last_name=u'Bradley', first_name=u'Antonio', s
sn=u'351-69-0455', occupation=u'Financial adviser', age=38), Row(last_n
ame=u'Lopez', first_name=u'Carol', ssn=u'499-61-5555', occupation=u'Lex
icographer', age=25), Row(last_name=u'Peterson', first_name=u'Courtne
y', ssn=u'721-32-0645', occupation=u'Psychiatrist', age=24), Row(last_n
ame=u'Boyd', first_name=u'Deborah', ssn=u'127-38-8755', occupation=u'Op
tometrist', age=31), Row(last_name=u'Kelly', first_name=u'Dr.', ssn=u'2
87-97-5503', occupation=u'Production assistant, radio', age=37), Row(la
st_name=u'Sandoval', first_name=u'Erika', ssn=u'629-17-0295', occupatio
n=u'Prison officer', age=24), Row(last_name=u'Evans', first_name=u'Jenn
ifer', ssn=u'473-03-8186', occupation=u'Cabin crew', age=34), Row(last_
name=u'Morris', first_name=u'Tami', ssn=u'708-32-0708', occupation=u'Sp
ort and exercise psychologist', age=42), Row(last_name=u'Mahoney', firs
t_name=u'Ashley', ssn=u'403-71-5709', occupation=u'Tax inspector', age=
3), Row(last_name=u'Kelly', first_name=u'Brittney', ssn=u'475-38-3019',
occupation=u'Haematologist', age=15), Row(last_name=u'Berger', first_na
me=u'Alexandra', ssn=u'418-72-3262', occupation=u'Health and safety ins
pector', age=32), Row(last_name=u'Mills', first_name=u'Louis', ssn=u'80
3-99-8901', occupation=u'Sports therapist', age=18), Row(last_name=u'Ar
mstrong', first_name=u'Megan', ssn=u'049-85-3941', occupation=u'Dramath
erapist', age=2), Row(last_name=u'Perez', first_name=u'Amy', ssn=u'099-
04-3826', occupation=u'Garment/textile technologist', age=14), Row(last
_name=u'Nunez', first_name=u'Shannon', ssn=u'112-21-8817', occupation=
u'Insurance underwriter', age=6), Row(last_name=u'Thomas', first_name=
u'Daniel', ssn=u'695-38-9638', occupation=u'Customer service manager',
 age=24), Row(last_name=u'Wiggins', first_name=u'Casey', ssn=u'649-67-3
465', occupation=u'Call centre manager', age=5), Row(last_name=u'Marti
n', first_name=u'Marissa', ssn=u'002-26-7246', occupation=u'Textile des
igner', age=28), Row(last_name=u'Adams', first_name=u'Joseph', ssn=u'24
8-08-7947', occupation=u'Tax inspector', age=34), Row(last_name=u'Roger
s', first_name=u'Michele', ssn=u'226-77-5583', occupation=u'English as
 a foreign language teacher', age=5), Row(last_name=u'Johnson', first_n
ame=u'Walter', ssn=u'539-13-8003', occupation=u'Chartered management ac
```

countant', age=20), Row(last_name=u'Hall', first_name=u'Renee', ssn=u'6
72-53-8203', occupation=u'Sound technician, broadcasting/film/video', a
ge=44), Row(last_name=u'Vega', first_name=u'Leslie', ssn=u'159-87-836
7', occupation=u'Immunologist', age=40), Row(last_name=u'Mills', first_
name=u'Kyle', ssn=u'520-32-6839', occupation=u'Patent examiner', age=4
6), Row(last_name=u'Huff', first_name=u'Michael', ssn=u'273-52-5066', o
ccupation=u'Quarry manager', age=29), Row(last_name=u'Reynolds', first_
name=u'Meagan', ssn=u'294-76-0379', occupation=u'Education officer, env
ironmental', age=18), Row(last_name=u'Wolfe', first_name=u'James', ssn=
u'593-57-0254', occupation=u'Psychologist, clinical', age=29), Row(last
_name=u'Love', first_name=u'Kristin', ssn=u'392-72-6814', occupation=
u'Race relations officer', age=11), Row(last_name=u'Watts', first_name=
u'Victoria', ssn=u'567-24-1502', occupation=u'Pharmacologist', age=43),
Row(last_name=u'Campos', first_name=u'James', ssn=u'149-59-8626', occup
ation=u'Adult guidance worker', age=22), Row(last_name=u'Maxwell', firs
t_name=u'Lauren', ssn=u'479-09-2372', occupation=u'Television productio
n assistant', age=34), Row(last_name=u'Woodward', first_name=u'Heathe
r', ssn=u'361-17-4136', occupation=u'Writer', age=26), Row(last_name=
u'Fernandez', first_name=u'Gregory', ssn=u'390-52-2752', occupation=u'F
orensic scientist', age=21), Row(last_name=u'Warren', first_name=u'Raym
ond', ssn=u'064-65-1288', occupation=u'Physiotherapist', age=44), Row(l
ast_name=u'Durham', first_name=u'John', ssn=u'248-24-2574', occupation=
u'Scientist, product/process development', age=22), Row(last_name=u'Whi
te', first_name=u'Dennis', ssn=u'454-03-3242', occupation=u'Minerals su
rveyor', age=43), Row(last_name=u'Porter', first_name=u'Richard', ssn=
u'779-81-8382', occupation=u'Garment/textile technologist', age=34), Ro
w(last_name=u'Bailey', first_name=u'Dominic', ssn=u'263-43-8648', occup
ation=u'Insurance underwriter', age=31), Row(last_name=u'Gould', first_
name=u'John', ssn=u'273-64-8976', occupation=u'Education officer, commu
nity', age=17), Row(last_name=u'George', first_name=u'Heidi', ssn=u'702
-16-3739', occupation=u'Accountant, chartered', age=45), Row(last_name=
u'Phillips', first_name=u'Angela', ssn=u'484-41-2877', occupation=u'Pla
nning and development surveyor', age=45), Row(last_name=u'Durham', firs
t_name=u'Brandi', ssn=u'415-67-5339', occupation=u'Accounting technicia
n', age=34), Row(last_name=u'Flores', first_name=u'Franklin', ssn=u'090
-21-2918', occupation=u'Musician', age=28), Row(last_name=u'Baker', fir
st_name=u'Frank', ssn=u'052-84-7029', occupation=u'Pharmacologist', age
=29), Row(last_name=u'Mcbride', first_name=u'Tara', ssn=u'772-92-2978',
occupation=u'Administrator, local government', age=42), Row(last_name=
u'Lewis', first_name=u'Lorraine', ssn=u'135-30-6367', occupation=u'Offi
ce manager', age=41), Row(last_name=u'Cox', first_name=u'Chelsea', ssn=
u'228-33-8234', occupation=u'Media buyer', age=31), Row(last_name=u'Ke
y', first_name=u'Jim', ssn=u'625-95-0479', occupation=u'Lobbyist', age=
39), Row(last_name=u'Gordon', first_name=u'Xavier', ssn=u'757-16-2546',
occupation=u'Amenity horticulturist', age=5), Row(last_name=u'Perez', f
irst_name=u'Julia', ssn=u'676-39-4923', occupation=u'Forensic scientis
t', age=7), Row(last_name=u'Welch', first_name=u'Crystal', ssn=u'660-88
-4361', occupation=u'Sport and exercise psychologist', age=37), Row(las
t_name=u'Thompson', first_name=u'Vanessa', ssn=u'222-39-8224', occupati
on=u'Warehouse manager', age=15), Row(last_name=u'Blackburn', first_nam
e=u'Karen', ssn=u'767-81-6609', occupation=u'Textile designer', age=4
3), Row(last_name=u'Boyd', first_name=u'Alison', ssn=u'441-13-9330', oc
cupation=u'Exhibition designer', age=29), Row(last_name=u'Wilson', firs
t_name=u'Jacob', ssn=u'662-16-4347', occupation=u'IT consultant', age=3
2), Row(last_name=u'Hull', first_name=u'Mark', ssn=u'436-09-7453', occu
pation=u'Photographer', age=3), Row(last_name=u'Colleen', first_name=
u'Mrs.', ssn=u'555-25-3192', occupation=u'Medical technical officer', a

ge=21), Row(last_name=u'Palmer', first_name=u'Robert', ssn=u'891-48-064
1', occupation=u'Psychologist, occupational', age=39), Row(last_name=
u'Lewis', first_name=u'Stephanie', ssn=u'096-15-4776', occupation=u'Sci
entist, product/process development', age=31), Row(last_name=u'George',
first_name=u'Eric', ssn=u'263-77-0976', occupation=u'Risk analyst', age
=25), Row(last_name=u'Rodriguez', first_name=u'Rebecca', ssn=u'210-95-3
436', occupation=u'Insurance underwriter', age=10), Row(last_name=u'All
en', first_name=u'Mark', ssn=u'328-44-3291', occupation=u'Journalist, n
ewspaper', age=11), Row(last_name=u'King', first_name=u'Caleb', ssn=u'3
58-11-8753', occupation=u'Senior tax professional/tax inspector', age=3
9), Row(last_name=u'Rodriguez', first_name=u'Aaron', ssn=u'633-26-005
0', occupation=u'IT consultant', age=2), Row(last_name=u'Lutz', first_n
ame=u'Lynn', ssn=u'311-52-3331', occupation=u'Marine scientist', age=2
4), Row(last_name=u'Duncan', first_name=u'Jesus', ssn=u'132-55-5090', o
ccupation=u'Financial risk analyst', age=8), Row(last_name=u'Horton', f
irst_name=u'Natalie', ssn=u'010-66-0981', occupation=u'Sport and exerci
se psychologist', age=3), Row(last_name=u'Copeland', first_name=u'Jerem
iah', ssn=u'317-40-0830', occupation=u'Contracting civil engineer', age
=14), Row(last_name=u'Lee', first_name=u'Jennifer', ssn=u'573-08-8621',
occupation=u'Oceanographer', age=33), Row(last_name=u'Evans', first_nam
e=u'Paul', ssn=u'295-37-9207', occupation=u'Insurance risk surveyor', a
ge=35), Row(last_name=u'Taylor', first_name=u'Kathryn', ssn=u'297-18-74
64', occupation=u'Paramedic', age=26), Row(last_name=u'Matthews', first
_name=u'Nicholas', ssn=u'414-81-3193', occupation=u'Librarian, public',
age=20), Row(last_name=u'Owens', first_name=u'Jason', ssn=u'636-76-126
6', occupation=u'Media buyer', age=15), Row(last_name=u'Gonzales', firs
t_name=u'Nicole', ssn=u'389-80-3753', occupation=u'Web designer', age=4
5), Row(last_name=u'Mcclure', first_name=u'Desiree', ssn=u'165-87-967
5', occupation=u'Corporate treasurer', age=21), Row(last_name=u'Mckenzi
e', first_name=u'Larry', ssn=u'897-09-9600', occupation=u'Editor, film/
video', age=10), Row(last_name=u'Case', first_name=u'Ryan', ssn=u'586-6
6-6045', occupation=u'Scientist, biomedical', age=38), Row(last_name=
u'Grant', first_name=u'Kelly', ssn=u'087-32-2169', occupation=u'Mining
 engineer', age=42), Row(last_name=u'Martinez', first_name=u'Troy', ssn
=u'251-51-1335', occupation=u'Therapist, speech and language', age=22),
Row(last_name=u'Mitchell', first_name=u'Bonnie', ssn=u'218-11-2543', oc
cupation=u'Production assistant, television', age=44), Row(last_name=
u'Leach', first_name=u'Cassidy', ssn=u'544-65-2828', occupation=u'Educa
tional psychologist', age=22), Row(last_name=u'Bridges', first_name=u'B
rian', ssn=u'577-79-6966', occupation=u'Television/film/video produce
r', age=6), Row(last_name=u'Santos', first_name=u'Cameron', ssn=u'167-8
2-5832', occupation=u'Commercial art gallery manager', age=14), Row(las
t_name=u'Schwartz', first_name=u'Brandon', ssn=u'063-03-7767', occupati
on=u'Event organiser', age=31), Row(last_name=u'Rodriguez', first_name=
u'Jeffrey', ssn=u'410-46-4436', occupation=u'Haematologist', age=45), R
ow(last_name=u'Foster', first_name=u'Rebecca', ssn=u'862-66-2237', occu
pation=u'Futures trader', age=31), Row(last_name=u'Roach', first_name=
u'Darius', ssn=u'434-29-3406', occupation=u'Tax inspector', age=19), Ro
w(last_name=u'Rogers', first_name=u'Joseph', ssn=u'051-10-8609', occupa
tion=u'Environmental education officer', age=23), Row(last_name=u'Vazqu
ez', first_name=u'Angel', ssn=u'371-59-8543', occupation=u'Medical phys
icist', age=46), Row(last_name=u'Johnson', first_name=u'Jessica', ssn=
u'563-17-6496', occupation=u'Legal executive', age=28), Row(last_name=
u'Mills', first_name=u'Gary', ssn=u'587-20-9041', occupation=u'Clinical
embryologist', age=35), Row(last_name=u'Watts', first_name=u'Robert', s
sn=u'276-07-8182', occupation=u'Financial trader', age=38), Row(last_na
me=u'Bradley', first_name=u'Julie', ssn=u'075-98-6978', occupation=u'Br

```
oadcast engineer', age=39), Row(last_name=u'Miller', first_name=u'Miche
lle', ssn=u'494-60-1400', occupation=u'Firefighter', age=26), Row(last_
name=u'Johnson', first_name=u'Amy', ssn=u'006-78-4009', occupation=u'Es
tate agent', age=45), Row(last_name=u'Norton', first_name=u'Christophe
r', ssn=u'580-14-6806', occupation=u'Therapist, occupational', age=40),
Row(last_name=u'Aguirre', first_name=u'Cristian', ssn=u'035-66-0540', o
ccupation=u'Surveyor, minerals', age=46), Row(last_name=u'Butler', firs
t_name=u'Colleen', ssn=u'358-91-1304', occupation=u'Quantity surveyor',
age=6), Row(last_name=u'Duke', first_name=u'Roger', ssn=u'897-42-8108',
occupation=u'Physiotherapist', age=38), Row(last_name=u'Schroeder', fir
st_name=u'Kelly', ssn=u'165-14-4145', occupation=u'Accountant, chartere
d management', age=39), Row(last_name=u'Bell', first_name=u'Christian',
ssn=u'889-64-6688', occupation=u'Adult nurse', age=36), Row(last_name=
u'Good', first_name=u'Kenneth', ssn=u'851-86-9928', occupation=u'Scient
ist, clinical (histocompatibility and immunogenetics)', age=27), Row(la
st_name=u'Cook', first_name=u'David', ssn=u'619-97-2652', occupation=
u'Retail merchandiser', age=15), Row(last_name=u'Gonzalez', first_name=
u'Emily', ssn=u'073-78-8261', occupation=u'Forest/woodland manager', ag
e=9), Row(last_name=u'Alvarado', first_name=u'Mitchell', ssn=u'207-21-0
183', occupation=u'Designer, graphic', age=21), Row(last_name=u'Munoz',
first_name=u'Kristin', ssn=u'109-56-8753', occupation=u'Sub', age=13),
 Row(last_name=u'Harvey', first_name=u'Eric', ssn=u'043-57-8449', occup
ation=u'Information systems manager', age=45), Row(last_name=u'Jimene
z', first_name=u'Julia', ssn=u'143-36-5276', occupation=u'Firefighter',
age=12), Row(last_name=u'James', first_name=u'Todd', ssn=u'151-18-745
0', occupation=u'Soil scientist', age=17), Row(last_name=u'Williamson',
first_name=u'Felicia', ssn=u'030-39-8799', occupation=u'Hydrologist', a
ge=27), Row(last_name=u'Pierce', first_name=u'Jonathan', ssn=u'222-65-9
192', occupation=u'Food technologist', age=44), Row(last_name=u'Rosari
o', first_name=u'Krystal', ssn=u'434-72-2348', occupation=u'Garment/tex
tile technologist', age=39), Row(last_name=u'Johnson', first_name=u'Ben
jamin', ssn=u'264-70-7605', occupation=u'Insurance claims handler', age
=0), Row(last_name=u'Michael', first_name=u'Dr.', ssn=u'663-65-7324', o
ccupation=u'Charity fundraiser', age=21), Row(last_name=u'Joyce', first
_name=u'Kevin', ssn=u'832-44-6634', occupation=u'Recycling officer', ag
e=44), Row(last_name=u'Matthews', first_name=u'Thomas', ssn=u'067-62-75
44', occupation=u'Environmental manager', age=22), Row(last_name=u'Leo
n', first_name=u'John', ssn=u'108-15-0252', occupation=u'Designer, exhi
bition/display', age=33), Row(last_name=u'Casey', first_name=u'Lisa', s
sn=u'059-18-3574', occupation=u'Mining engineer', age=26), Row(last_nam
e=u'Roberts', first_name=u'Thomas', ssn=u'168-40-1409', occupation=u'Mu
seum/gallery curator', age=28), Row(last_name=u'Davis', first_name=u'Ja
cob', ssn=u'727-87-2811', occupation=u'Best boy', age=26), Row(last_nam
e=u'Parsons', first_name=u'Julia', ssn=u'237-25-7914', occupation=u'Cor
porate investment banker', age=44), Row(last_name=u'Schneider', first_n
ame=u'Nicholas', ssn=u'185-65-8245', occupation=u'Public librarian', ag
e=1), Row(last_name=u'Novak', first_name=u'Christopher', ssn=u'590-41-5
573', occupation=u'Librarian, public', age=7), Row(last_name=u'George',
first_name=u'Gabrielle', ssn=u'487-80-8213', occupation=u'Research scie
ntist (medical)', age=18), Row(last_name=u'Brown', first_name=u'Richar
d', ssn=u'766-97-3117', occupation=u'Water quality scientist', age=46),
Row(last_name=u'Ramirez', first_name=u'Travis', ssn=u'677-74-7778', occ
upation=u'Engineer, drilling', age=11), Row(last_name=u'Harrison', firs
t_name=u'Jenna', ssn=u'633-27-7745', occupation=u'Telecommunications re
searcher', age=22), Row(last_name=u'Horton', first_name=u'Kelly', ssn=
u'765-69-3888', occupation=u'Government social research officer', age=4
5), Row(last_name=u'Coleman', first_name=u'Zachary', ssn=u'368-98-710
```

6', occupation=u'Scientist, water quality', age=29), Row(last_name=u'Fr
ost', first_name=u'Robert', ssn=u'633-88-8219', occupation=u'Geneticis
t, molecular', age=0), Row(last_name=u'Olson', first_name=u'Matthew', s
sn=u'212-90-9667', occupation=u'Planning and development surveyor', age
=17), Row(last_name=u'Bond', first_name=u'Charlene', ssn=u'230-08-096
4', occupation=u'Corporate treasurer', age=4), Row(last_name=u'Hines',
 first_name=u'Sean', ssn=u'347-10-3366', occupation=u'Planning and deve
lopment surveyor', age=2), Row(last_name=u'Flores', first_name=u'Alex',
ssn=u'301-58-0213', occupation=u'Charity fundraiser', age=28), Row(last
_name=u'Castillo', first_name=u'Victoria', ssn=u'422-03-8227', occupati
on=u'Conference centre manager', age=28), Row(last_name=u'Black', first
_name=u'William', ssn=u'433-48-7671', occupation=u'Immunologist', age=2
0), Row(last_name=u'Johnson', first_name=u'Jason', ssn=u'613-82-1597',
 occupation=u'Information officer', age=19), Row(last_name=u'Jones', fi
rst_name=u'Michael', ssn=u'183-45-8066', occupation=u'Surveyor, minin
g', age=19), Row(last_name=u'Guerra', first_name=u'Victoria', ssn=u'323
-88-7831', occupation=u'Office manager', age=15), Row(last_name=u'Cunni
ngham', first_name=u'Danielle', ssn=u'341-58-2176', occupation=u'Databa
se administrator', age=30), Row(last_name=u'Robinson', first_name=u'Dan
iel', ssn=u'214-92-2018', occupation=u'Technical author', age=8), Row(l
ast_name=u'Rowe', first_name=u'James', ssn=u'404-56-4652', occupation=
u'Ship broker', age=34), Row(last_name=u'Burgess', first_name=u'Sue', s
sn=u'028-18-3538', occupation=u'Actuary', age=7), Row(last_name=u'Hardi
ng', first_name=u'Ashley', ssn=u'353-15-5146', occupation=u'Chartered a
ccountant', age=4), Row(last_name=u'Rollins', first_name=u'Kevin', ssn=
u'834-14-2783', occupation=u'Water quality scientist', age=36), Row(las
t_name=u'Wright', first_name=u'Troy', ssn=u'788-02-5866', occupation=
u'Pension scheme manager', age=13), Row(last_name=u'Gonzalez', first_na
me=u'Brian', ssn=u'496-97-6122', occupation=u'Leisure centre manager',
 age=7), Row(last_name=u'Ward', first_name=u'Dwayne', ssn=u'607-81-274
9', occupation=u'Cabin crew', age=26), Row(last_name=u'Edwards', first_
name=u'Daniel', ssn=u'838-60-2931', occupation=u'Engineer, manufacturin
g systems', age=24), Row(last_name=u'Ford', first_name=u'Monica', ssn=
u'141-54-4231', occupation=u'Clinical research associate', age=6), Row
(last_name=u'Garrett', first_name=u'Amanda', ssn=u'724-95-6759', occupa
tion=u'Mudlogger', age=15), Row(last_name=u'Martin', first_name=u'Natas
ha', ssn=u'688-53-9085', occupation=u'Location manager', age=13), Row(l
ast_name=u'Ross', first_name=u'Michele', ssn=u'889-48-0997', occupation
=u'Ophthalmologist', age=26), Row(last_name=u'Houston', first_name=u'Ca
itlin', ssn=u'675-15-8916', occupation=u'Personnel officer', age=43), R
ow(last_name=u'Sanchez', first_name=u'Elizabeth', ssn=u'782-87-4531', o
ccupation=u'Geneticist, molecular', age=14), Row(last_name=u'Lopez', fi
rst_name=u'Jessica', ssn=u'587-65-9872', occupation=u'Clinical biochemi
st', age=43), Row(last_name=u'Taylor', first_name=u'Matthew', ssn=u'001
-24-5516', occupation=u'Psychiatric nurse', age=2), Row(last_name=u'Hoo
per', first_name=u'William', ssn=u'233-91-9045', occupation=u'Energy ma
nager', age=6), Row(last_name=u'Hoffman', first_name=u'John', ssn=u'881
-84-2403', occupation=u'Museum/gallery exhibitions officer', age=30), R
ow(last_name=u'Thornton', first_name=u'Tina', ssn=u'729-72-5966', occup
ation=u'Sport and exercise psychologist', age=33), Row(last_name=u'Stan
ley', first_name=u'Kelly', ssn=u'488-32-7085', occupation=u'Financial p
lanner', age=14), Row(last_name=u'Hines', first_name=u'Joseph', ssn=u'5
28-20-5677', occupation=u'Private music teacher', age=9), Row(last_name
=u'Thompson', first_name=u'Shane', ssn=u'338-11-1453', occupation=u'Inv
estment analyst', age=41), Row(last_name=u'Barber', first_name=u'Dillo
n', ssn=u'457-57-9361', occupation=u'Recruitment consultant', age=4), R
ow(last_name=u'Brown', first_name=u'Alyssa', ssn=u'582-38-9387', occupa

tion=u'Graphic designer', age=20), Row(last_name=u'Crawford', first_nam
e=u'Yvonne', ssn=u'735-01-2026', occupation=u'Personal assistant', age=
40), Row(last_name=u'Terrell', first_name=u'Daniel', ssn=u'520-74-974
1', occupation=u'Buyer, retail', age=22), Row(last_name=u'Schmitt', fir
st_name=u'James', ssn=u'290-14-5541', occupation=u'Armed forces logisti
cs/support/administrative officer', age=11), Row(last_name=u'Mckenzie',
first_name=u'Molly', ssn=u'836-29-4083', occupation=u'Management consul
tant', age=9), Row(last_name=u'Payne', first_name=u'William', ssn=u'846
-29-5360', occupation=u'Pharmacist, community', age=3), Row(last_name=
u'Davis', first_name=u'Virginia', ssn=u'150-93-0656', occupation=u'Tech
nical sales engineer', age=10), Row(last_name=u'Smith', first_name=u'Co
urtney', ssn=u'724-30-6581', occupation=u'Chiropractor', age=8), Row(la
st_name=u'Rodriguez', first_name=u'Sarah', ssn=u'671-37-8443', occupati
on=u'Gaffer', age=25), Row(last_name=u'Stark', first_name=u'Brenda', ss
n=u'160-33-8962', occupation=u'Tree surgeon', age=25), Row(last_name=
u'Johnson', first_name=u'Jessica', ssn=u'249-15-0782', occupation=u'Tex
tile designer', age=37), Row(last_name=u'Clarke', first_name=u'Jacqueli
ne', ssn=u'117-55-0563', occupation=u'Ship broker', age=3), Row(last_na
me=u'Kelli', first_name=u'Mrs.', ssn=u'745-51-5001', occupation=u'Biome
dical engineer', age=20), Row(last_name=u'Bishop', first_name=u'Manue
l', ssn=u'417-50-2061', occupation=u'Computer games developer', age=3
0), Row(last_name=u'Peterson', first_name=u'Jeremiah', ssn=u'498-97-378
8', occupation=u'Embryologist, clinical', age=2), Row(last_name=u'Yan
g', first_name=u'Alan', ssn=u'178-44-2079', occupation=u'Ecologist', ag
e=7), Row(last_name=u'Coleman', first_name=u'Jonathan', ssn=u'618-27-83
43', occupation=u'Conference centre manager', age=9), Row(last_name=u'R
ichardson', first_name=u'Ethan', ssn=u'172-11-3013', occupation=u'Illus
trator', age=14), Row(last_name=u'Williams', first_name=u'Theresa', ssn
=u'719-35-6021', occupation=u'Technical author', age=16), Row(last_name
=u'Johnson', first_name=u'Christopher', ssn=u'896-39-6243', occupation=
u'Immunologist', age=16), Row(last_name=u'Beasley', first_name=u'Elizab
eth', ssn=u'408-16-7047', occupation=u'Social research officer, governm
ent', age=42), Row(last_name=u'Smith', first_name=u'Brittany', ssn=u'10
3-34-4821', occupation=u'Sales executive', age=14), Row(last_name=u'Wil
son', first_name=u'Jessica', ssn=u'476-81-6375', occupation=u'Newspaper
journalist', age=16), Row(last_name=u'French', first_name=u'Jesus', ssn
=u'327-66-4065', occupation=u'Agricultural engineer', age=45), Row(last
_name=u'Dixon', first_name=u'Megan', ssn=u'053-58-2932', occupation=u'T
echnical brewer', age=23), Row(last_name=u'Garcia', first_name=u'Gabrie
l', ssn=u'154-17-8930', occupation=u'Metallurgist', age=19), Row(last_n
ame=u'Coleman', first_name=u'Albert', ssn=u'047-34-1894', occupation=
u'Therapist, music', age=19), Row(last_name=u'Thompson', first_name=u'N
icole', ssn=u'740-35-5073', occupation=u'Interior and spatial designe
r', age=4), Row(last_name=u'Torres', first_name=u'Katie', ssn=u'638-05-
9578', occupation=u'Television camera operator', age=7), Row(last_name=
u'Jacobs', first_name=u'Olivia', ssn=u'668-88-8599', occupation=u'Engin
eer, mining', age=37), Row(last_name=u'Woodward', first_name=u'Nathan',
ssn=u'363-87-7735', occupation=u'Scientist, forensic', age=17), Row(las
t_name=u'Weber', first_name=u'Melody', ssn=u'737-93-3153', occupation=
u'Community development worker', age=7), Row(last_name=u'Walters', firs
t_name=u'Jonathan', ssn=u'229-87-6308', occupation=u'Location manager',
age=33), Row(last_name=u'Holland', first_name=u'Christy', ssn=u'290-82-
3045', occupation=u'Astronomer', age=13), Row(last_name=u'Stone', first
_name=u'Samantha', ssn=u'751-89-7943', occupation=u'IT consultant', age
=21), Row(last_name=u'Green', first_name=u'Angel', ssn=u'326-12-2653',
occupation=u'Data processing manager', age=28), Row(last_name=u'Mcdona
ld', first_name=u'Gregory', ssn=u'747-92-7298', occupation=u'Film/video

editor', age=27), Row(last_name=u'Salinas', first_name=u'Paul', ssn=u'6
11-13-0736', occupation=u'Production assistant, television', age=17), R
ow(last_name=u'Perez', first_name=u'Erin', ssn=u'823-51-8422', occupati
on=u'Production engineer', age=33), Row(last_name=u'Marshall', first_na
me=u'Jonathan', ssn=u'545-04-8538', occupation=u'Archaeologist', age=
9), Row(last_name=u'Smith', first_name=u'Kimberly', ssn=u'210-14-4506',
occupation=u'Environmental health practitioner', age=35), Row(last_name
=u'Morrow', first_name=u'Charles', ssn=u'283-88-6257', occupation=u'Pro
grammer, multimedia', age=9), Row(last_name=u'Clark', first_name=u'Eri
k', ssn=u'139-79-2907', occupation=u'Scientist, research (life science
s)', age=20), Row(last_name=u'Barker', first_name=u'Dakota', ssn=u'532-
73-0002', occupation=u'Art therapist', age=41), Row(last_name=u'Norto
n', first_name=u'Ashley', ssn=u'766-24-8966', occupation=u'Child psycho
therapist', age=42), Row(last_name=u'Riddle', first_name=u'William', ss
n=u'403-56-8785', occupation=u'Network engineer', age=7), Row(last_name
=u'Jones', first_name=u'Vickie', ssn=u'250-67-1749', occupation=u'Horti
culturist, amenity', age=37), Row(last_name=u'Fisher', first_name=u'Nan
cy', ssn=u'820-55-7976', occupation=u'Surveyor, commercial/residentia
l', age=22), Row(last_name=u'Fernandez', first_name=u'Kelly', ssn=u'265
-92-3771', occupation=u'Video editor', age=16), Row(last_name=u'Becke
r', first_name=u'Alexa', ssn=u'172-90-5227', occupation=u'Interpreter',
age=44), Row(last_name=u'Garcia', first_name=u'Cynthia', ssn=u'500-30-5
165', occupation=u'Designer, television/film set', age=16), Row(last_na
me=u'Wallace', first_name=u'Stephanie', ssn=u'718-51-8122', occupation=
u'Audiological scientist', age=11), Row(last_name=u'Timothy', first_nam
e=u'Mr.', ssn=u'454-76-9272', occupation=u'Public librarian', age=0), R
ow(last_name=u'Key', first_name=u'Barbara', ssn=u'821-36-3832', occupat
ion=u'Computer games developer', age=34), Row(last_name=u'Hill', first_
name=u'Amanda', ssn=u'572-52-0388', occupation=u'Minerals surveyor', ag
e=9), Row(last_name=u'Carey', first_name=u'Martin', ssn=u'457-15-8882',
occupation=u'Television production assistant', age=40), Row(last_name=
u'Murray', first_name=u'Elizabeth', ssn=u'054-88-8396', occupation=u'Me
dical laboratory scientific officer', age=21), Row(last_name=u'Mcgee',
 first_name=u'Jessica', ssn=u'037-01-2015', occupation=u'Network engine
er', age=27), Row(last_name=u'Simmons', first_name=u'Michael', ssn=u'67
3-46-1497', occupation=u'Trading standards officer', age=33), Row(last_
name=u'Hogan', first_name=u'Ashley', ssn=u'703-09-3728', occupation=u'P
rogrammer, multimedia', age=22), Row(last_name=u'Woods', first_name=u'A
nthony', ssn=u'755-26-3427', occupation=u'Medical secretary', age=19),
 Row(last_name=u'Bradley', first_name=u'Yvette', ssn=u'436-56-6643', oc
cupation=u'Commercial/residential surveyor', age=19), Row(last_name=u'R
odriguez', first_name=u'Rose', ssn=u'686-17-7340', occupation=u'Town pl
anner', age=5), Row(last_name=u'Williams', first_name=u'Andrew', ssn=
u'650-42-9736', occupation=u'Health promotion specialist', age=28), Row
(last_name=u'Kramer', first_name=u'Barbara', ssn=u'052-88-5532', occupa
tion=u'Soil scientist', age=22), Row(last_name=u'Velazquez', first_name
=u'Nicholas', ssn=u'078-60-1807', occupation=u'Colour technologist', ag
e=26), Row(last_name=u'Nielsen', first_name=u'Zachary', ssn=u'188-89-42
02', occupation=u'Research scientist (life sciences)', age=7), Row(last
_name=u'Shepard', first_name=u'Kelly', ssn=u'020-18-9707', occupation=
u'Colour technologist', age=30), Row(last_name=u'Willis', first_name=
u'Joseph', ssn=u'211-73-3953', occupation=u'Teacher, primary school', a
ge=36), Row(last_name=u'Graham', first_name=u'Bethany', ssn=u'005-47-86
68', occupation=u'Special effects artist', age=4), Row(last_name=u'Wats
on', first_name=u'Andrew', ssn=u'565-55-3879', occupation=u'Public rela
tions account executive', age=24), Row(last_name=u'Green', first_name=
u'Jonathan', ssn=u'497-91-1681', occupation=u'Ambulance person', age=2

9), Row(last_name=u'Adams', first_name=u'Cynthia', ssn=u'506-84-1814',
 occupation=u'Magazine features editor', age=12), Row(last_name=u'Smit
h', first_name=u'Olivia', ssn=u'203-35-9255', occupation=u'Telecommunic
ations researcher', age=25), Row(last_name=u'Krueger', first_name=u'Joh
n', ssn=u'378-52-5286', occupation=u'Therapist, art', age=33), Row(last
_name=u'Morales', first_name=u'Evelyn', ssn=u'683-86-2651', occupation=
u'Public relations officer', age=20), Row(last_name=u'Ramirez', first_n
ame=u'Deborah', ssn=u'702-48-7793', occupation=u'Engineer, automotive',
age=8), Row(last_name=u'Porter', first_name=u'Ryan', ssn=u'126-07-664
3', occupation=u'Research officer, trade union', age=22), Row(last_name
=u'Gardner', first_name=u'Kelly', ssn=u'400-79-1473', occupation=u'Lear
ning mentor', age=24), Row(last_name=u'Barton', first_name=u'Justin', s
sn=u'730-60-3348', occupation=u'Psychotherapist', age=33), Row(last_nam
e=u'Lambert', first_name=u'Michelle', ssn=u'175-14-3767', occupation=
u'Contracting civil engineer', age=4), Row(last_name=u'Fields', first_n
ame=u'Kelsey', ssn=u'709-58-9762', occupation=u'Research scientist (phy
sical sciences)', age=35), Row(last_name=u'Kim', first_name=u'Michael',
ssn=u'660-37-5605', occupation=u'Insurance claims handler', age=18), Ro
w(last_name=u'Thomas', first_name=u'Mr.', ssn=u'462-18-6201', occupatio
n=u'Ergonomist', age=41), Row(last_name=u'Page', first_name=u'Claudia',
ssn=u'082-91-3791', occupation=u'Accountant, chartered certified', age=
27), Row(last_name=u'White', first_name=u'Sara', ssn=u'096-83-9829', oc
cupation=u'Advertising art director', age=38), Row(last_name=u'Matthew
s', first_name=u'William', ssn=u'691-54-8930', occupation=u'Financial a
dviser', age=44), Row(last_name=u'Summers', first_name=u'Daniel', ssn=
u'342-10-0244', occupation=u'Web designer', age=19), Row(last_name=u'Br
own', first_name=u'Kevin', ssn=u'053-67-5558', occupation=u'Accommodati
on manager', age=3), Row(last_name=u'Kelly', first_name=u'Michael', ssn
=u'015-79-1959', occupation=u'Editorial assistant', age=8), Row(last_na
me=u'Ford', first_name=u'Steve', ssn=u'449-81-0650', occupation=u'Energ
y engineer', age=8), Row(last_name=u'Fuller', first_name=u'Hannah', ssn
=u'526-11-5291', occupation=u'Environmental manager', age=39), Row(last
_name=u'Gallagher', first_name=u'Jennifer', ssn=u'514-66-4585', occupat
ion=u'Tax inspector', age=42), Row(last_name=u'Wallace', first_name=u'M
elissa', ssn=u'768-31-1862', occupation=u'Sales promotion account execu
tive', age=11), Row(last_name=u'Rasmussen', first_name=u'Latoya', ssn=
u'118-51-4984', occupation=u'Insurance account manager', age=38), Row(l
ast_name=u'Vaughn', first_name=u'Jamie', ssn=u'763-81-1298', occupation
=u'Games developer', age=13), Row(last_name=u'Collins', first_name=u'Je
anne', ssn=u'788-55-6338', occupation=u'Early years teacher', age=36),
 Row(last_name=u'Willis', first_name=u'Richard', ssn=u'592-83-6373', oc
cupation=u'Tree surgeon', age=35), Row(last_name=u'Marshall', first_nam
e=u'Theresa', ssn=u'613-10-1156', occupation=u'Designer, furniture', ag
e=19), Row(last_name=u'Cochran', first_name=u'Tommy', ssn=u'305-02-110
8', occupation=u'Dentist', age=17), Row(last_name=u'Phillips', first_na
me=u'Paula', ssn=u'143-35-4520', occupation=u'Research officer, governm
ent', age=22), Row(last_name=u'Landry', first_name=u'James', ssn=u'567-
58-5095', occupation=u'Surveyor, building control', age=4), Row(last_na
me=u'Joseph', first_name=u'Dr.', ssn=u'624-34-4208', occupation=u'Plann
ing and development surveyor', age=23), Row(last_name=u'Stevenson', fir
st_name=u'Angela', ssn=u'228-29-6832', occupation=u'Risk analyst', age=
41), Row(last_name=u'Davis', first_name=u'Brittany', ssn=u'180-30-037
2', occupation=u'Engineer, technical sales', age=45), Row(last_name=u'C
lark', first_name=u'Alan', ssn=u'050-01-8908', occupation=u'Fashion des
igner', age=14), Row(last_name=u'Boyd', first_name=u'Heidi', ssn=u'028-
03-7237', occupation=u'Agricultural engineer', age=30), Row(last_name=
u'Daniel', first_name=u'Brenda', ssn=u'584-72-3906', occupation=u'Armed

forces technical officer', age=34), Row(last_name=u'Russell', first_nam
e=u'Amy', ssn=u'162-96-9141', occupation=u'Engineer, structural', age=
5), Row(last_name=u'Barker', first_name=u'Madeline', ssn=u'590-37-136
3', occupation=u'Therapist, horticultural', age=30), Row(last_name=u'Fe
rnandez', first_name=u'Deborah', ssn=u'038-27-9421', occupation=u'Surve
yor, hydrographic', age=23), Row(last_name=u'Johnson', first_name=u'Mic
hael', ssn=u'326-74-8762', occupation=u'Sports administrator', age=9),
 Row(last_name=u'Dennis', first_name=u'Katrina', ssn=u'810-66-4528', oc
cupation=u'IT technical support officer', age=9), Row(last_name=u'Thomp
son', first_name=u'Ryan', ssn=u'113-31-2667', occupation=u'Tourism offi
cer', age=0), Row(last_name=u'Lewis', first_name=u'Richard', ssn=u'319-
21-2567', occupation=u'Chemist, analytical', age=29), Row(last_name=u'W
atson', first_name=u'Traci', ssn=u'178-63-1565', occupation=u'Catering
 manager', age=20), Row(last_name=u'Davis', first_name=u'Benjamin', ssn
=u'031-86-4105', occupation=u'Engineer, manufacturing', age=39), Row(la
st_name=u'Tucker', first_name=u'Thomas', ssn=u'161-85-8277', occupation
=u'Interior and spatial designer', age=10), Row(last_name=u'Brock', fir
st_name=u'Anthony', ssn=u'156-92-2430', occupation=u'Programmer, system
s', age=24), Row(last_name=u'Jackson', first_name=u'Andrea', ssn=u'824-
04-8610', occupation=u'English as a foreign language teacher', age=12),
Row(last_name=u'Jones', first_name=u'Diana', ssn=u'436-70-5618', occupa
tion=u'Company secretary', age=39), Row(last_name=u'Rodriguez', first_n
ame=u'Rachael', ssn=u'503-69-7201', occupation=u'Civil engineer, contra
cting', age=2), Row(last_name=u'Johnston', first_name=u'Daniel', ssn=
u'161-97-4162', occupation=u'Dancer', age=19), Row(last_name=u'Hardin',
first_name=u'Elizabeth', ssn=u'848-21-7982', occupation=u'Risk manage
r', age=32), Row(last_name=u'Smith', first_name=u'Brooke', ssn=u'706-65
-2798', occupation=u"Nurse, children's", age=37), Row(last_name=u'Mcdon
ald', first_name=u'Tyler', ssn=u'160-09-9140', occupation=u'Aid worke
r', age=18), Row(last_name=u'Mercado', first_name=u'Richard', ssn=u'530
-30-0095', occupation=u'Museum education officer', age=21), Row(last_na
me=u'Boone', first_name=u'Sarah', ssn=u'663-68-4674', occupation=u'Prod
uct designer', age=10), Row(last_name=u'Thomas', first_name=u'Crystal',
ssn=u'261-84-7648', occupation=u'Video editor', age=45), Row(last_name=
u'Roberts', first_name=u'Dylan', ssn=u'387-29-9392', occupation=u'Desig
ner, multimedia', age=4), Row(last_name=u'Garcia', first_name=u'Thoma
s', ssn=u'251-22-4817', occupation=u'Outdoor activities/education manag
er', age=29), Row(last_name=u'Marks', first_name=u'Ian', ssn=u'895-99-8
212', occupation=u'Academic librarian', age=10), Row(last_name=u'Evan
s', first_name=u'Sean', ssn=u'517-68-0711', occupation=u'Interior and s
patial designer', age=13), Row(last_name=u'Maxwell', first_name=u'Hanna
h', ssn=u'592-25-3981', occupation=u'Programmer, systems', age=34), Row
(last_name=u'Stafford', first_name=u'Robert', ssn=u'216-72-1451', occup
ation=u'Estate agent', age=0), Row(last_name=u'Smith', first_name=u'Sar
ah', ssn=u'809-70-5607', occupation=u'Early years teacher', age=14), Ro
w(last_name=u'Scott', first_name=u'Victor', ssn=u'252-83-2263', occupat
ion=u'Trading standards officer', age=28), Row(last_name=u'Taylor', fir
st_name=u'David', ssn=u'626-61-6272', occupation=u"Barrister's clerk",
 age=0), Row(last_name=u'Klein', first_name=u'Colleen', ssn=u'885-82-52
22', occupation=u'Television floor manager', age=4), Row(last_name=u'Ar
nold', first_name=u'Kevin', ssn=u'363-51-2895', occupation=u'Health ser
vice manager', age=14), Row(last_name=u'Lynn', first_name=u'Jessica', s
sn=u'353-12-9753', occupation=u'Conservation officer, historic building
s', age=22), Row(last_name=u'Lucero', first_name=u'Micheal', ssn=u'113-
44-4054', occupation=u'Television/film/video producer', age=26), Row(la
st_name=u'Clark', first_name=u'Andrew', ssn=u'160-40-0561', occupation=
u'Air traffic controller', age=31), Row(last_name=u'Smith', first_name=

u'Leslie', ssn=u'668-45-1213', occupation=u'Chartered accountant', age=
31), Row(last_name=u'Berry', first_name=u'Jennifer', ssn=u'816-94-594
0', occupation=u'Lexicographer', age=20), Row(last_name=u'Guzman', firs
t_name=u'Natasha', ssn=u'692-18-7034', occupation=u'Race relations offi
cer', age=9), Row(last_name=u'Paul', first_name=u'Lori', ssn=u'627-92-7
170', occupation=u'Therapeutic radiographer', age=17), Row(last_name=
u'Schaefer', first_name=u'Jamie', ssn=u'043-92-2227', occupation=u'Engi
neer, automotive', age=43), Row(last_name=u'Martinez', first_name=u'Rob
in', ssn=u'040-16-2542', occupation=u'Product designer', age=0), Row(la
st_name=u'Cobb', first_name=u'Tony', ssn=u'791-70-3110', occupation=u'D
octor, hospital', age=18), Row(last_name=u'Schneider', first_name=u'Bre
nda', ssn=u'589-67-2746', occupation=u'Teacher, special educational nee
ds', age=37), Row(last_name=u'Coleman', first_name=u'Jane', ssn=u'496-8
6-9276', occupation=u'Social worker', age=4), Row(last_name=u'Nicholso
n', first_name=u'Jacqueline', ssn=u'372-77-8058', occupation=u'Occupati
onal therapist', age=32), Row(last_name=u'Price', first_name=u'Corey',
 ssn=u'841-33-7133', occupation=u'Records manager', age=25), Row(last_n
ame=u'Harris', first_name=u'Linda', ssn=u'277-03-8762', occupation=u'Su
rveyor, land/geomatics', age=29), Row(last_name=u'Luna', first_name=u'R
obert', ssn=u'695-07-2347', occupation=u'Local government officer', age
=5), Row(last_name=u'Smith', first_name=u'Kathryn', ssn=u'049-99-2260',
occupation=u'Quantity surveyor', age=14), Row(last_name=u'Long', first_
name=u'Michael', ssn=u'350-17-1045', occupation=u'Youth worker', age=3
5), Row(last_name=u'Andrade', first_name=u'Amy', ssn=u'829-25-7984', oc
cupation=u'Drilling engineer', age=36), Row(last_name=u'Williams', firs
t_name=u'William', ssn=u'127-64-5574', occupation=u'Travel agency manag
er', age=26), Row(last_name=u'Gray', first_name=u'Leslie', ssn=u'789-83
-2286', occupation=u'Insurance claims handler', age=31), Row(last_name=
u'Tucker', first_name=u'Michael', ssn=u'793-96-3103', occupation=u'Spor
ts coach', age=43), Row(last_name=u'Rodriguez', first_name=u'Michael',
 ssn=u'572-83-4180', occupation=u'Therapist, horticultural', age=14), R
ow(last_name=u'Harrison', first_name=u'Annette', ssn=u'274-96-6979', oc
cupation=u'Scientist, biomedical', age=40), Row(last_name=u'Scott', fir
st_name=u'Paul', ssn=u'759-50-8389', occupation=u'Civil Service adminis
trator', age=2), Row(last_name=u'Adrian', first_name=u'Mr.', ssn=u'841-
19-4993', occupation=u'Osteopath', age=1), Row(last_name=u'Flynn', firs
t_name=u'Jennifer', ssn=u'672-38-2029', occupation=u'Land/geomatics sur
veyor', age=35), Row(last_name=u'Arellano', first_name=u'Pamela', ssn=
u'053-48-9855', occupation=u'Health and safety inspector', age=24), Row
(last_name=u'Johns', first_name=u'Bradley', ssn=u'213-71-7361', occupat
ion=u'Futures trader', age=6), Row(last_name=u'Nguyen', first_name=u'De
siree', ssn=u'447-66-6188', occupation=u'Scientist, water quality', age
=28), Row(last_name=u'Mercer', first_name=u'Adam', ssn=u'369-24-7779',
 occupation=u'Firefighter', age=15), Row(last_name=u'Martinez', first_n
ame=u'Cindy', ssn=u'295-24-1041', occupation=u'Designer, ceramics/potte
ry', age=20), Row(last_name=u'Williams', first_name=u'Jessica', ssn=u'8
49-79-9790', occupation=u'Artist', age=32), Row(last_name=u'Torres', fi
rst_name=u'Jasmine', ssn=u'466-83-6282', occupation=u'Nutritional thera
pist', age=30), Row(last_name=u'Romero', first_name=u'Abigail', ssn=u'1
94-58-3344', occupation=u'Designer, interior/spatial', age=27), Row(las
t_name=u'Evans', first_name=u'Gregory', ssn=u'273-10-0298', occupation=
u'Illustrator', age=29), Row(last_name=u'Smith', first_name=u'Samuel',
 ssn=u'619-66-6766', occupation=u'Advertising account executive', age=
1), Row(last_name=u'Price', first_name=u'Alan', ssn=u'038-13-9203', occ
upation=u'Medical technical officer', age=25), Row(last_name=u'Lee', fi
rst_name=u'Kelly', ssn=u'085-86-0441', occupation=u'Forensic scientis
t', age=3), Row(last_name=u'Mclaughlin', first_name=u'Aaron', ssn=u'555

-45-8619', occupation=u'Geochemist', age=20), Row(last_name=u'Madden',
 first_name=u'Kathleen', ssn=u'535-53-2726', occupation=u'Engineer, mai
ntenance', age=40), Row(last_name=u'Zavala', first_name=u'Michael', ssn
=u'190-34-1581', occupation=u'Learning mentor', age=12), Row(last_name=
u'Gonzalez', first_name=u'Andrew', ssn=u'148-54-9498', occupation=u'Mag
azine journalist', age=41), Row(last_name=u'Holmes', first_name=u'Thoma
s', ssn=u'303-45-6028', occupation=u'Teaching laboratory technician', a
ge=2), Row(last_name=u'Hurst', first_name=u'Laurie', ssn=u'854-46-
8536', occupation=u'Surveyor, land/geomatics', age=41), Row(last_name=
u'Brown', first_name=u'Robert', ssn=u'486-54-1506', occupation=u'Charte
red public finance accountant', age=26), Row(last_name=u'Lara', first_n
ame=u'Kevin', ssn=u'280-02-6091', occupation=u'Automotive engineer', ag
e=17), Row(last_name=u'Lopez', first_name=u'Donald', ssn=u'710-94-293
3', occupation=u'Engineer, maintenance (IT)', age=3), Row(last_name=u'C
hambers', first_name=u'Dawn', ssn=u'679-41-9669', occupation=u'Enginee
r, land', age=26), Row(last_name=u'Morton', first_name=u'Karen', ssn=
u'395-59-4087', occupation=u'Designer, ceramics/pottery', age=9), Row(l
ast_name=u'Salazar', first_name=u'Danielle', ssn=u'507-93-2466', occupa
tion=u'Engineer, manufacturing systems', age=25), Row(last_name=u'Jone
s', first_name=u'Austin', ssn=u'734-70-0615', occupation=u'Financial ri
sk analyst', age=13), Row(last_name=u'Weber', first_name=u'James', ssn=
u'896-75-9561', occupation=u'Engineer, electrical', age=3), Row(last_na
me=u'Washington', first_name=u'Philip', ssn=u'674-43-8950', occupation=
u'Social researcher', age=20), Row(last_name=u'Combs', first_name=u'Jac
ob', ssn=u'552-72-9451', occupation=u'Higher education lecturer', age=3
1), Row(last_name=u'Knight', first_name=u'Matthew', ssn=u'104-18-5597',
 occupation=u'Information systems manager', age=45), Row(last_name=u'Al
len', first_name=u'Brett', ssn=u'545-69-0495', occupation=u'Surveyor, q
uantity', age=8), Row(last_name=u'Reed', first_name=u'Hannah', ssn=u'38
5-85-0768', occupation=u'Archaeologist', age=11), Row(last_name=u'Cisne
ros', first_name=u'Sydney', ssn=u'389-88-7992', occupation=u'Archaeolog
ist', age=44), Row(last_name=u'Ward', first_name=u'Tyler', ssn=u'593-17
-7079', occupation=u'Land', age=23), Row(last_name=u'Henderson', first_
name=u'Tracy', ssn=u'543-32-1172', occupation=u'Loss adjuster, chartere
d', age=33), Row(last_name=u'Murray', first_name=u'Brittany', ssn=u'126
-43-0244', occupation=u'Land', age=9), Row(last_name=u'Giles', first_na
me=u'Jessica', ssn=u'395-43-1230', occupation=u'Fine artist', age=1), R
ow(last_name=u'Morgan', first_name=u'Scott', ssn=u'751-81-0926', occupa
tion=u'Toxicologist', age=7), Row(last_name=u'Miller', first_name=u'Bri
an', ssn=u'562-74-5426', occupation=u'Health promotion specialist', age
=8), Row(last_name=u'Jones', first_name=u'Bobby', ssn=u'319-94-9000', o
ccupation=u'Engineer, production', age=16), Row(last_name=u'Frye', firs
t_name=u'Allison', ssn=u'382-99-9693', occupation=u'Petroleum enginee
r', age=19), Row(last_name=u'Stewart', first_name=u'Corey', ssn=u'232-5
2-0228', occupation=u'Police officer', age=45), Row(last_name=u'Everet
t', first_name=u'Luis', ssn=u'416-75-2519', occupation=u'Animal technol
ogist', age=27), Row(last_name=u'Wilkins', first_name=u'Shawna', ssn=
u'807-47-7396', occupation=u'Animal technologist', age=30), Row(last_na
me=u'Livingston', first_name=u'Brandon', ssn=u'102-99-5017', occupation
=u'Health service manager', age=14), Row(last_name=u'Hall', first_name=
u'Kimberly', ssn=u'673-20-6362', occupation=u'Historic buildings inspec
tor/conservation officer', age=8), Row(last_name=u'Lopez', first_name=
u'Vanessa', ssn=u'860-95-5743', occupation=u'Education officer, museu
m', age=19), Row(last_name=u'Howell', first_name=u'Benjamin', ssn=u'655
-97-3735', occupation=u'Engineer, civil (consulting)', age=38), Row(las
t_name=u'Myers', first_name=u'Jeffrey', ssn=u'229-38-0774', occupation=
u'Secretary, company', age=39), Row(last_name=u'Johnson', first_name=

```
u'Sydney', ssn=u'831-53-7676', occupation=u'Field trials officer', age=
27), Row(last_name=u'Harper', first_name=u'Steven', ssn=u'417-45-2194',
 occupation=u'Merchandiser, retail', age=42), Row(last_name=u'Duarte',
 first_name=u'Amber', ssn=u'583-45-6226', occupation=u'Ecologist', age=
10), Row(last_name=u'Cabrera', first_name=u'Gary', ssn=u'761-66-8236',
 occupation=u'Electrical engineer', age=23), Row(last_name=u'Snyder', f
irst_name=u'Lisa', ssn=u'310-95-0302', occupation=u'Accommodation manag
er', age=27), Row(last_name=u'Briggs', first_name=u'Amber', ssn=u'765-7
2-0879', occupation=u'Buyer, retail', age=1), Row(last_name=u'Robbins',
 first_name=u'Scott', ssn=u'312-84-8842', occupation=u'Designer, exhibi
tion/display', age=15), Row(last_name=u'Wilkins', first_name=u'Michae
l', ssn=u'355-51-1545', occupation=u'Research scientist (medical)', age
=7), Row(last_name=u'Harris', first_name=u'Jeffrey', ssn=u'093-43-295
5', occupation=u'Colour technologist', age=30), Row(last_name=u'Walke
r', first_name=u'Katie', ssn=u'610-54-7475', occupation=u'Event organis
er', age=45), Row(last_name=u'Ruiz', first_name=u'Brian', ssn=u'758-79-
0211', occupation=u'Journalist, newspaper', age=23), Row(last_name=u'Mi
ller', first_name=u'Anthony', ssn=u'550-66-8766', occupation=u'Cabin cr
ew', age=32), Row(last_name=u'Ray', first_name=u'Laura', ssn=u'519-70-5
246', occupation=u'Warehouse manager', age=42), Row(last_name=u'Payne',
 first_name=u'Casey', ssn=u'387-83-5576', occupation=u'Community develo
pment worker', age=33), Row(last_name=u'Neal', first_name=u'Vanessa', s
sn=u'720-88-3148', occupation=u'Editor, commissioning', age=26), Row(la
st_name=u'Brooks', first_name=u'Timothy', ssn=u'505-23-2619', occupatio
n=u'Optician, dispensing', age=26), Row(last_name=u'Adkins', first_name
=u'James', ssn=u'526-52-9732', occupation=u'Veterinary surgeon', age=3
3), Row(last_name=u'Harris', first_name=u'Sarah', ssn=u'805-75-8482', o
ccupation=u'Visual merchandiser', age=34), Row(last_name=u'Thornton', f
irst_name=u'Andrew', ssn=u'608-35-4126', occupation=u'Colour technologi
st', age=42), Row(last_name=u'Jones', first_name=u'Leah', ssn=u'680-88-
0072', occupation=u'Mudlogger', age=31), Row(last_name=u'Gray', first_n
ame=u'Shane', ssn=u'265-02-2800', occupation=u'Housing manager/office
r', age=7), Row(last_name=u'Carter', first_name=u'Barbara', ssn=u'535-8
8-4637', occupation=u'Buyer, industrial', age=38), Row(last_name=u'Lov
e', first_name=u'Amy', ssn=u'558-13-6963', occupation=u'Administrator,
 local government', age=3), Row(last_name=u'Bradley', first_name=u'Nich
olas', ssn=u'619-14-5596', occupation=u'Scientist, physiological', age=
6), Row(last_name=u'Morgan', first_name=u'Shawn', ssn=u'372-45-2043', o
ccupation=u'Surveyor, building', age=40), Row(last_name=u'Smith', first
_name=u'Jennifer', ssn=u'520-75-3206', occupation=u'Historic buildings
 inspector/conservation officer', age=10), Row(last_name=u'French', fir
st_name=u'Karen', ssn=u'435-84-6808', occupation=u'Sports development o
fficer', age=46), Row(last_name=u'Nguyen', first_name=u'Lauren', ssn=
u'006-55-4079', occupation=u'Acupuncturist', age=1), Row(last_name=u'Sc
hwartz', first_name=u'Michelle', ssn=u'442-62-5799', occupation=u'Sound
 technician, broadcasting/film/video', age=0), Row(last_name=u'Lucas',
 first_name=u'Douglas', ssn=u'781-51-9842', occupation=u'Theatre manage
r', age=27), Row(last_name=u'Herman', first_name=u'Gary', ssn=u'048-02-
5497', occupation=u'Pharmacist, community', age=3), Row(last_name=u'Bra
y', first_name=u'Christopher', ssn=u'159-10-8737', occupation=u'Metallu
rgist', age=9), Row(last_name=u'Howard', first_name=u'William', ssn=u'0
60-31-1323', occupation=u'Exercise physiologist', age=6), Row(last_name
=u'Gilbert', first_name=u'Sandra', ssn=u'495-57-7539', occupation=u'Inf
ormation systems manager', age=40), Row(last_name=u'Freeman', first_nam
e=u'Amanda', ssn=u'750-22-7752', occupation=u'Theatre manager', age=3
8), Row(last_name=u'Robles', first_name=u'Holly', ssn=u'106-45-1294', o
ccupation=u'Building services engineer', age=1), Row(last_name=u'Lee',
```

 first_name=u'David', ssn=u'129-20-0705', occupation=u'Industrial/produ
ct designer', age=7), Row(last_name=u'Allen', first_name=u'Jennifer', s
sn=u'683-33-6603', occupation=u'Engineer, technical sales', age=28), Ro
w(last_name=u'Knight', first_name=u'Tanya', ssn=u'293-47-3361', occupat
ion=u'Rural practice surveyor', age=10), Row(last_name=u'Garcia', first
_name=u'Richard', ssn=u'511-53-6288', occupation=u'Stage manager', age=
11), Row(last_name=u'Case', first_name=u'Andrew', ssn=u'580-32-5712', o
ccupation=u'Private music teacher', age=4), Row(last_name=u'Johnson', f
irst_name=u'Christine', ssn=u'242-24-9866', occupation=u'Waste manageme
nt officer', age=30), Row(last_name=u'Mullen', first_name=u'Jordan', ss
n=u'615-70-1311', occupation=u'Production assistant, television', age=2
1), Row(last_name=u'Harris', first_name=u'Sarah', ssn=u'689-91-2858', o
ccupation=u'Higher education lecturer', age=17), Row(last_name=u'Pere
z', first_name=u'Thomas', ssn=u'631-13-7676', occupation=u'Physiotherap
ist', age=33), Row(last_name=u'Lucas', first_name=u'Sergio', ssn=u'772-
40-6854', occupation=u'Ship broker', age=46), Row(last_name=u'Gonzale
z', first_name=u'Juan', ssn=u'094-91-7769', occupation=u'Electronics en
gineer', age=21), Row(last_name=u'Walton', first_name=u'Christina', ssn
=u'783-92-4175', occupation=u'Community arts worker', age=19), Row(last
_name=u'Kaiser', first_name=u'Stephanie', ssn=u'693-20-9026', occupatio
n=u'Museum/gallery curator', age=6), Row(last_name=u'Cochran', first_na
me=u'Sheila', ssn=u'576-31-5722', occupation=u'Scientist, clinical (his
tocompatibility and immunogenetics)', age=9), Row(last_name=u'Levy', fi
rst_name=u'Michelle', ssn=u'740-17-4508', occupation=u'Biomedical scien
tist', age=13), Row(last_name=u'Susan', first_name=u'Mrs.', ssn=u'428-4
4-6872', occupation=u'Firefighter', age=17), Row(last_name=u'Ramos', fi
rst_name=u'Eric', ssn=u'672-20-8025', occupation=u'Teacher, secondary s
chool', age=11), Row(last_name=u'Olson', first_name=u'Katie', ssn=u'407
-46-3898', occupation=u'Solicitor', age=45), Row(last_name=u'Richards',
 first_name=u'Adrian', ssn=u'318-82-9797', occupation=u'Pharmacologis
t', age=19), Row(last_name=u'Hines', first_name=u'Kristine', ssn=u'595-
12-5598', occupation=u'Archivist', age=28), Row(last_name=u'Rodriguez',
 first_name=u'Patricia', ssn=u'071-68-1444', occupation=u'Art therapis
t', age=1), Row(last_name=u'Lee', first_name=u'Pamela', ssn=u'696-59-01
32', occupation=u'Accommodation manager', age=3), Row(last_name=u'Hanso
n', first_name=u'Richard', ssn=u'422-59-3375', occupation=u'Health serv
ice manager', age=27), Row(last_name=u'Lynch', first_name=u'Kevin', ssn
=u'458-83-4767', occupation=u'Outdoor activities/education manager', ag
e=18), Row(last_name=u'Odom', first_name=u'Harold', ssn=u'496-98-2439',
 occupation=u'Geologist, engineering', age=39), Row(last_name=u'Gonzale
z', first_name=u'Deborah', ssn=u'150-35-5469', occupation=u'Licensed co
nveyancer', age=30), Row(last_name=u'Dennis', first_name=u'Richard', ss
n=u'390-04-0565', occupation=u'Scientist, marine', age=15), Row(last_na
me=u'Scott', first_name=u'Daniel', ssn=u'576-35-2979', occupation=u'Fin
ancial manager', age=25), Row(last_name=u'Clark', first_name=u'Barbar
a', ssn=u'658-50-5745', occupation=u'Diagnostic radiographer', age=37),
 Row(last_name=u'Allen', first_name=u'Donald', ssn=u'065-26-7295', occu
pation=u'Teacher, music', age=35), Row(last_name=u'Arnold', first_name=
u'Mia', ssn=u'195-27-4019', occupation=u'Product designer', age=39), Ro
w(last_name=u'Pena', first_name=u'Teresa', ssn=u'371-93-1893', occupati
on=u'Tree surgeon', age=20), Row(last_name=u'Calderon', first_name=u'Ch
arles', ssn=u'069-50-0871', occupation=u'Investment analyst', age=45),
 Row(last_name=u'Collins', first_name=u'Randy', ssn=u'871-59-4668', occ
upation=u'Catering manager', age=21), Row(last_name=u'Hughes', first_na
me=u'Kristin', ssn=u'445-03-0748', occupation=u'Operational investment
 banker', age=22), Row(last_name=u'Gray', first_name=u'Richard', ssn=
u'608-07-1863', occupation=u'Amenity horticulturist', age=0), Row(last_

```
name=u'Hanna', first_name=u'Bianca', ssn=u'012-83-9402', occupation=u'I
nsurance broker', age=0), Row(last_name=u'Jones', first_name=u'Thomas',
 ssn=u'016-51-7271', occupation=u'Patent examiner', age=21), Row(last_n
ame=u'Walton', first_name=u'Andres', ssn=u'013-50-1928', occupation=u'E
nvironmental manager', age=44), Row(last_name=u'Cochran', first_name=
u'Scott', ssn=u'463-59-5335', occupation=u'Dramatherapist', age=3), Row
(last_name=u'James', first_name=u'Sally', ssn=u'386-53-4905', occupatio
n=u'Equality and diversity officer', age=35), Row(last_name=u'Flores',
 first_name=u'Jennifer', ssn=u'775-29-2541', occupation=u'Environmental
 education officer', age=4), Row(last_name=u'Krueger', first_name=u'Bri
ana', ssn=u'540-33-0830', occupation=u'Academic librarian', age=36), Ro
w(last_name=u'Lee', first_name=u'Betty', ssn=u'190-39-9570', occupation
=u'Surveyor, building control', age=21), Row(last_name=u'Johnson', firs
t_name=u'Gregory', ssn=u'573-39-8752', occupation=u'Engineer, mining',
 age=29), Row(last_name=u'Johnson', first_name=u'Cory', ssn=u'678-78-53
47', occupation=u'Clinical biochemist', age=20), Row(last_name=u'Gardne
r', first_name=u'Jason', ssn=u'178-30-3494', occupation=u'Therapist, mu
sic', age=26), Row(last_name=u'Mitchell', first_name=u'Michael', ssn=
u'321-90-8320', occupation=u'Teacher, special educational needs', age=
3), Row(last_name=u'Schmidt', first_name=u'Melissa', ssn=u'190-52-086
4', occupation=u'Information systems manager', age=13), Row(last_name=
u'Gill', first_name=u'Debra', ssn=u'308-81-5452', occupation=u'Administ
rator, Civil Service', age=46), Row(last_name=u'Steele', first_name=u'T
yler', ssn=u'313-63-9020', occupation=u'Catering manager', age=7), Row
(last_name=u'Cook', first_name=u'Susan', ssn=u'711-88-0577', occupation
=u'Control and instrumentation engineer', age=19), Row(last_name=u'Fowl
er', first_name=u'Joshua', ssn=u'625-75-7838', occupation=u'Dispensing
 optician', age=45), Row(last_name=u'Smith', first_name=u'Christopher',
 ssn=u'737-79-2349', occupation=u'Applications developer', age=7), Row
(last_name=u'Kelley', first_name=u'Courtney', ssn=u'358-19-0313', occup
ation=u'Hospital doctor', age=15), Row(last_name=u'Mccullough', first_n
ame=u'Sandra', ssn=u'391-35-9478', occupation=u'Hydrogeologist', age=1
8), Row(last_name=u'Ramirez', first_name=u'Matthew', ssn=u'821-58-880
8', occupation=u'Midwife', age=17), Row(last_name=u'Bridges', first_nam
e=u'Robert', ssn=u'620-74-5333', occupation=u'Insurance claims handle
r', age=44), Row(last_name=u'Reeves', first_name=u'Duane', ssn=u'706-25
-6584', occupation=u'Fashion designer', age=35), Row(last_name=u'Gordo
n', first_name=u'Stephanie', ssn=u'663-14-5658', occupation=u'Land', ag
e=18), Row(last_name=u'Smith', first_name=u'Andre', ssn=u'287-30-7107',
 occupation=u'Health service manager', age=26), Row(last_name=u'Jones',
 first_name=u'Christopher', ssn=u'368-99-5367', occupation=u'Tax advise
r', age=24), Row(last_name=u'Pitts', first_name=u'Lynn', ssn=u'050-97-8
589', occupation=u'Orthoptist', age=37), Row(last_name=u'Kim', first_na
me=u'Aaron', ssn=u'622-09-9301', occupation=u'Film/video editor', age=1
1), Row(last_name=u'Waters', first_name=u'Robert', ssn=u'532-59-6236',
 occupation=u'Pilot, airline', age=38), Row(last_name=u'Avila', first_n
ame=u'Elizabeth', ssn=u'643-59-9537', occupation=u'Lobbyist', age=25),
 Row(last_name=u'Gray', first_name=u'Brittany', ssn=u'121-10-1987', occ
upation=u'Health and safety adviser', age=8), Row(last_name=u'Hernande
z', first_name=u'Douglas', ssn=u'075-72-7141', occupation=u'Teacher, se
condary school', age=10), Row(last_name=u'Wheeler', first_name=u'Mary',
 ssn=u'887-06-4489', occupation=u'Tour manager', age=22), Row(last_name
=u'Sanford', first_name=u'Savannah', ssn=u'536-69-6417', occupation=u'J
ournalist, magazine', age=2), Row(last_name=u'Thomas', first_name=u'Dav
id', ssn=u'521-25-1443', occupation=u'Engineer, manufacturing', age=1
4), Row(last_name=u'Jones', first_name=u'Derek', ssn=u'743-55-0071', oc
cupation=u'Veterinary surgeon', age=22), Row(last_name=u'Owens', first_
```

name=u'John', ssn=u'617-49-0765', occupation=u'Pensions consultant', ag
e=6), Row(last_name=u'Black', first_name=u'Melanie', ssn=u'275-80-637
8', occupation=u'Psychologist, educational', age=3), Row(last_name=u'Co
peland', first_name=u'Sarah', ssn=u'450-82-6587', occupation=u'Private
 music teacher', age=29), Row(last_name=u'Lopez', first_name=u'Emily',
 ssn=u'153-06-8389', occupation=u'Optician, dispensing', age=8), Row(la
st_name=u'Lewis', first_name=u'Jessica', ssn=u'742-72-2957', occupation
=u'Petroleum engineer', age=1), Row(last_name=u'Silva', first_name=u'Va
nessa', ssn=u'374-65-2575', occupation=u'Chiropractor', age=25), Row(la
st_name=u'Scott', first_name=u'Christopher', ssn=u'588-01-8231', occupa
tion=u'Forensic scientist', age=26), Row(last_name=u'Warren', first_nam
e=u'Ashley', ssn=u'696-50-9949', occupation=u'Corporate investment bank
er', age=29), Row(last_name=u'Buchanan', first_name=u'Gary', ssn=u'040-
67-7312', occupation=u'Sports coach', age=12), Row(last_name=u'Johnso
n', first_name=u'Lynn', ssn=u'397-62-2113', occupation=u'IT technical s
upport officer', age=6), Row(last_name=u'Brown', first_name=u'Joshua',
 ssn=u'176-48-9342', occupation=u'Therapeutic radiographer', age=23), R
ow(last_name=u'Watkins', first_name=u'Michael', ssn=u'809-16-5154', occ
upation=u'Marine scientist', age=28), Row(last_name=u'Kevin', first_nam
e=u'Mr.', ssn=u'711-91-9101', occupation=u'Secretary/administrator', ag
e=34), Row(last_name=u'Adams', first_name=u'Samuel', ssn=u'048-35-537
1', occupation=u'Therapist, horticultural', age=1), Row(last_name=u'Rob
erts', first_name=u'Rick', ssn=u'382-16-6732', occupation=u'Administrat
or, local government', age=40), Row(last_name=u'Smith', first_name=u'Vi
ctoria', ssn=u'651-37-2473', occupation=u'Clinical research associate',
 age=31), Row(last_name=u'White', first_name=u'Robert', ssn=u'304-99-69
17', occupation=u'Community development worker', age=3), Row(last_name=
u'Martinez', first_name=u'Erin', ssn=u'806-64-2496', occupation=u'Theat
re stage manager', age=36), Row(last_name=u'Proctor', first_name=u'Megh
an', ssn=u'773-72-4200', occupation=u'Teacher, English as a foreign lan
guage', age=0), Row(last_name=u'Hodge', first_name=u'Madison', ssn=u'21
8-70-2502', occupation=u'Architect', age=38), Row(last_name=u'Jimenez',
 first_name=u'Deborah', ssn=u'490-09-3413', occupation=u'Production des
igner, theatre/television/film', age=37), Row(last_name=u'Perry', first
_name=u'Kylie', ssn=u'400-51-8586', occupation=u'Air broker', age=5), R
ow(last_name=u'Ponce', first_name=u'Sergio', ssn=u'192-29-9271', occupa
tion=u'Animal nutritionist', age=12), Row(last_name=u'Soto', first_name
=u'Riley', ssn=u'409-82-3336', occupation=u'Accounting technician', age
=17), Row(last_name=u'Collier', first_name=u'Justin', ssn=u'838-63-468
0', occupation=u'Historic buildings inspector/conservation officer', ag
e=24), Row(last_name=u'Mendoza', first_name=u'Sandy', ssn=u'261-40-711
8', occupation=u'Field trials officer', age=45), Row(last_name=u'Taylo
r', first_name=u'Larry', ssn=u'756-57-3013', occupation=u'Legal secreta
ry', age=33), Row(last_name=u'Friedman', first_name=u'James', ssn=u'661
-50-3183', occupation=u'Engineer, water', age=26), Row(last_name=u'Ric
e', first_name=u'Adam', ssn=u'699-02-6633', occupation=u'Nature conserv
ation officer', age=42), Row(last_name=u'Dillon', first_name=u'Nichola
s', ssn=u'577-66-4146', occupation=u'Publishing copy', age=11), Row(las
t_name=u'Petersen', first_name=u'Joshua', ssn=u'838-90-0999', occupatio
n=u'Community pharmacist', age=34), Row(last_name=u'Walker', first_name
=u'Sarah', ssn=u'073-34-8340', occupation=u'Sport and exercise psycholo
gist', age=31), Row(last_name=u'Stewart', first_name=u'Taylor', ssn=u'6
80-74-3189', occupation=u'Research scientist (physical sciences)', age=
40), Row(last_name=u'Ramos', first_name=u'David', ssn=u'115-99-9252', o
ccupation=u'Computer games developer', age=40), Row(last_name=u'Stewar
t', first_name=u'Alexis', ssn=u'199-60-0085', occupation=u'Cartographe
r', age=23), Row(last_name=u'David', first_name=u'Mark', ssn=u'760-12-2

340', occupation=u'Designer, furniture', age=27), Row(last_name=u'Williamson', first_name=u'Sharon', ssn=u'382-03-7791', occupation=u'Occupational psychologist', age=31), Row(last_name=u'Lindsey', first_name=u'Maria', ssn=u'679-75-5174', occupation=u'Teacher, secondary school', age=8), Row(last_name=u'Smith', first_name=u'Carrie', ssn=u'283-31-5041', occupation=u'Surveyor, commercial/residential', age=38), Row(last_name=u'Doyle', first_name=u'Alicia', ssn=u'387-30-2633', occupation=u'Adult guidance worker', age=10), Row(last_name=u'Jimenez', first_name=u'Nicole', ssn=u'211-24-0406', occupation=u'Social research officer, government', age=32), Row(last_name=u'Lopez', first_name=u'Tina', ssn=u'025-87-9829', occupation=u'Warden/ranger', age=18), Row(last_name=u'Reyes', first_name=u'Charles', ssn=u'052-83-4063', occupation=u'Telecommunications researcher', age=14), Row(last_name=u'Anderson', first_name=u'Nicole', ssn=u'479-65-1369', occupation=u'Advertising copywriter', age=19), Row(last_name=u'Jacobs', first_name=u'Calvin', ssn=u'092-64-7059', occupation=u'Community development worker', age=31), Row(last_name=u'Mckee', first_name=u'Patricia', ssn=u'219-64-5159', occupation=u'Scientist, research (medical)', age=46), Row(last_name=u'Weaver', first_name=u'Emily', ssn=u'793-38-9931', occupation=u'Cytogeneticist', age=28), Row(last_name=u'Golden', first_name=u'Charlotte', ssn=u'814-33-1663', occupation=u'Fitness centre manager', age=20), Row(last_name=u'Sosa', first_name=u'Sheila', ssn=u'126-84-8217', occupation=u'Conservator, furniture', age=0), Row(last_name=u'Green', first_name=u'Adam', ssn=u'831-74-7410', occupation=u'Therapist, horticultural', age=16), Row(last_name=u'Ritter', first_name=u'Hannah', ssn=u'457-91-1375', occupation=u'Education officer, community', age=20), Row(last_name=u'Lindsey', first_name=u'Mrs.', ssn=u'409-67-7118', occupation=u'Seismic interpreter', age=16), Row(last_name=u'Patrick', first_name=u'Karen', ssn=u'627-99-8325', occupation=u'Equities trader', age=19), Row(last_name=u'Carpenter', first_name=u'Ryan', ssn=u'556-88-3621', occupation=u'Social research officer, government', age=8), Row(last_name=u'Baker', first_name=u'Bryan', ssn=u'602-50-7169', occupation=u'Immunologist', age=46), Row(last_name=u'Young', first_name=u'Matthew', ssn=u'396-76-7488', occupation=u'Food technologist', age=42), Row(last_name=u'Dixon', first_name=u'Melissa', ssn=u'644-63-6977', occupation=u'Press sub', age=20), Row(last_name=u'Wells', first_name=u'Terry', ssn=u'711-36-6810', occupation=u'Journalist, newspaper', age=36), Row(last_name=u'Johnson', first_name=u'Susan', ssn=u'442-08-9432', occupation=u'Media buyer', age=19), Row(last_name=u'Daniels', first_name=u'Suzanne', ssn=u'731-73-4978', occupation=u'Social researcher', age=10), Row(last_name=u'Newman', first_name=u'Ryan', ssn=u'328-94-0423', occupation=u'Medical illustrator', age=36), Row(last_name=u'Sharp', first_name=u'Amanda', ssn=u'335-54-2519', occupation=u'Leisure centre manager', age=20), Row(last_name=u'Ryan', first_name=u'Joseph', ssn=u'201-23-8853', occupation=u'Chartered management accountant', age=18), Row(last_name=u'Kent', first_name=u'Edwin', ssn=u'021-56-6400', occupation=u'Administrator, sports', age=24), Row(last_name=u'Ward', first_name=u'Steven', ssn=u'487-59-1064', occupation=u'Health promotion specialist', age=27), Row(last_name=u'Hunter', first_name=u'Jacqueline', ssn=u'873-25-1621', occupation=u'Sound technician, broadcasting/film/video', age=25), Row(last_name=u'Li', first_name=u'Frank', ssn=u'659-34-5395', occupation=u'Medical secretary', age=21), Row(last_name=u'Shaw', first_name=u'Michael', ssn=u'072-55-2887', occupation=u'Site engineer', age=31), Row(last_name=u'Arnold', first_name=u'Christopher', ssn=u'542-82-8240', occupation=u'IT trainer', age=44), Row(last_name=u'Thomas', first_name=u'Eric', ssn=u'443-76-6485', occupation=u'Diplomatic Services operational officer', age=38), Row(last_name=u'Robinson', first_name=u'Cheryl', ssn=u'660-09-3172', occupation=u'Scientist, rese

arch (medical)', age=22), Row(last_name=u'Miller', first_name=u'David',
 ssn=u'315-75-2307', occupation=u'Cartographer', age=19), Row(last_name
=u'Jacobson', first_name=u'Lisa', ssn=u'771-89-0161', occupation=u'Cust
omer service manager', age=31), Row(last_name=u'Hernandez', first_name=
u'Jessica', ssn=u'524-44-4846', occupation=u'Insurance broker', age=1
4), Row(last_name=u'Martin', first_name=u'William', ssn=u'056-42-7035',
 occupation=u'Environmental consultant', age=14), Row(last_name=u'Jacks
on', first_name=u'Courtney', ssn=u'490-92-7237', occupation=u'Textile d
esigner', age=39), Row(last_name=u'Parker', first_name=u'Kevin', ssn=
u'094-26-7611', occupation=u'Arts development officer', age=18), Row(la
st_name=u'Spence', first_name=u'Jennifer', ssn=u'063-71-8067', occupati
on=u'Geochemist', age=36), Row(last_name=u'Mcclure', first_name=u'Pau
l', ssn=u'789-94-6632', occupation=u'Press sub', age=23), Row(last_name
=u'Osborn', first_name=u'Evan', ssn=u'004-15-1779', occupation=u'Academ
ic librarian', age=12), Row(last_name=u'Greene', first_name=u'Jacquelin
e', ssn=u'816-01-3287', occupation=u'Records manager', age=13), Row(las
t_name=u'Barnes', first_name=u'Carrie', ssn=u'509-34-8502', occupation=
u'Therapist, drama', age=33), Row(last_name=u'Kelly', first_name=u'Darr
ell', ssn=u'834-34-6472', occupation=u'Accountant, chartered public fin
ance', age=27), Row(last_name=u'Ruiz', first_name=u'Clinton', ssn=u'155
-46-9013', occupation=u'Electronics engineer', age=32), Row(last_name=
u'Alexander', first_name=u'Sheila', ssn=u'416-05-9800', occupation=u'On
cologist', age=25), Row(last_name=u'Williams', first_name=u'Jacquelin
e', ssn=u'472-39-2328', occupation=u'Loss adjuster, chartered', age=2
3), Row(last_name=u'Coleman', first_name=u'Ashley', ssn=u'368-93-8941',
 occupation=u'International aid/development worker', age=44), Row(last_
name=u'Ramirez', first_name=u'Stephanie', ssn=u'556-77-4996', occupatio
n=u'Journalist, newspaper', age=18), Row(last_name=u'Beasley', first_na
me=u'Sharon', ssn=u'504-62-5082', occupation=u'Audiological scientist',
 age=14), Row(last_name=u'Smith', first_name=u'Sylvia', ssn=u'238-49-11
39', occupation=u'Environmental education officer', age=7), Row(last_na
me=u'Jones', first_name=u'Carol', ssn=u'318-58-8311', occupation=u'Food
 technologist', age=22), Row(last_name=u'Wood', first_name=u'Emily', ss
n=u'803-84-5244', occupation=u'Architect', age=0), Row(last_name=u'Smit
h', first_name=u'Eric', ssn=u'138-16-5150', occupation=u'Teacher, adult
 education', age=19), Row(last_name=u'Farrell', first_name=u'Paula', ss
n=u'574-99-6613', occupation=u'Chartered management accountant', age=3
6), Row(last_name=u'Harper', first_name=u'Joseph', ssn=u'557-70-7045',
 occupation=u'Scientist, audiological', age=12), Row(last_name=u'Berr
y', first_name=u'Robert', ssn=u'412-61-6551', occupation=u'Adult nurs
e', age=29), Row(last_name=u'Brown', first_name=u'Patrick', ssn=u'015-4
7-0379', occupation=u'Scientist, clinical (histocompatibility and immun
ogenetics)', age=22), Row(last_name=u'Susan', first_name=u'Dr.', ssn=
u'888-68-9130', occupation=u'Surveyor, hydrographic', age=24), Row(last
_name=u'Jones', first_name=u'Colin', ssn=u'893-40-6491', occupation=u'D
iplomatic Services operational officer', age=12), Row(last_name=u'Zunig
a', first_name=u'Sherry', ssn=u'844-98-5090', occupation=u'Editor, fil
m/video', age=28), Row(last_name=u'Calhoun', first_name=u'James', ssn=
u'467-34-7484', occupation=u'Civil Service administrator', age=41), Row
(last_name=u'Harvey', first_name=u'Carol', ssn=u'029-20-8088', occupati
on=u'Advertising copywriter', age=40), Row(last_name=u'Adams', first_na
me=u'Tyler', ssn=u'393-67-7898', occupation=u'Food technologist', age=3
1), Row(last_name=u'Gutierrez', first_name=u'Hannah', ssn=u'019-11-999
9', occupation=u'Exhibitions officer, museum/gallery', age=31), Row(las
t_name=u'Harris', first_name=u'Ashley', ssn=u'423-55-9217', occupation=
u'Interior and spatial designer', age=15), Row(last_name=u'Gallegos', f
irst_name=u'Jennifer', ssn=u'667-14-4352', occupation=u'Special educati

onal needs teacher', age=4), Row(last_name=u'Pruitt', first_name=u'Kell
y', ssn=u'413-05-0141', occupation=u'Civil engineer, consulting', age=1
3), Row(last_name=u'Arias', first_name=u'Edwin', ssn=u'409-69-8487', oc
cupation=u'Metallurgist', age=34), Row(last_name=u'Reeves', first_name=
u'Kenneth', ssn=u'188-92-2969', occupation=u'Commissioning editor', age
=5), Row(last_name=u'Watson', first_name=u'Deborah', ssn=u'188-04-381
3', occupation=u'Call centre manager', age=22), Row(last_name=u'Conle
y', first_name=u'Sean', ssn=u'629-52-0870', occupation=u'Accountant, ch
artered public finance', age=45), Row(last_name=u'Walton', first_name=
u'Elizabeth', ssn=u'807-15-8641', occupation=u'Designer, textile', age=
37), Row(last_name=u'Jones', first_name=u'Kristopher', ssn=u'768-40-667
2', occupation=u'Trade union research officer', age=8), Row(last_name=
u'Sanchez', first_name=u'Jennifer', ssn=u'205-52-1176', occupation=u'Pa
ssenger transport manager', age=10), Row(last_name=u'Martinez', first_n
ame=u'Kathryn', ssn=u'772-17-5191', occupation=u'Therapist, speech and
 language', age=5), Row(last_name=u'Rodriguez', first_name=u'Amanda', s
sn=u'187-80-6438', occupation=u'Sports development officer', age=18), R
ow(last_name=u'White', first_name=u'Michael', ssn=u'197-54-7053', occup
ation=u'Community development worker', age=2), Row(last_name=u'Clark',
 first_name=u'Jeffrey', ssn=u'309-82-4726', occupation=u'Early years te
acher', age=19), Row(last_name=u'Brown', first_name=u'Vanessa', ssn=u'4
31-73-4254', occupation=u'Senior tax professional/tax inspector', age=1
7), Row(last_name=u'Lynch', first_name=u'Sabrina', ssn=u'414-27-9494',
 occupation=u'Therapist, speech and language', age=23), Row(last_name=
u'Park', first_name=u'Jeremy', ssn=u'053-42-4600', occupation=u'Bonds t
rader', age=6), Row(last_name=u'Anderson', first_name=u'Lauren', ssn=
u'609-74-7336', occupation=u'Fashion designer', age=35), Row(last_name=
u'Williams', first_name=u'Gerald', ssn=u'329-76-1136', occupation=u'Hea
lth and safety adviser', age=27), Row(last_name=u'Byrd', first_name=u'T
anya', ssn=u'885-59-0047', occupation=u'Scientific laboratory technicia
n', age=11), Row(last_name=u'Benjamin', first_name=u'Dennis', ssn=u'364
-22-0353', occupation=u'Training and development officer', age=4), Row
(last_name=u'Stokes', first_name=u'Maria', ssn=u'001-26-3657', occupati
on=u'Training and development officer', age=4), Row(last_name=u'Duran',
 first_name=u'Clinton', ssn=u'628-78-1439', occupation=u'Therapist, occ
upational', age=46), Row(last_name=u'Evans', first_name=u'Jessica', ssn
=u'091-66-9241', occupation=u'Osteopath', age=1), Row(last_name=u'Bal
l', first_name=u'Annette', ssn=u'402-94-1583', occupation=u'Production
 engineer', age=41), Row(last_name=u'Hess', first_name=u'Kristin', ssn=
u'815-80-0553', occupation=u'Programmer, applications', age=5), Row(las
t_name=u'White', first_name=u'Crystal', ssn=u'712-92-9591', occupation=
u'Science writer', age=4), Row(last_name=u'Brewer', first_name=u'Ashle
y', ssn=u'214-30-3470', occupation=u'Proofreader', age=23), Row(last_na
me=u'Compton', first_name=u'Tamara', ssn=u'736-60-0208', occupation=u'D
esigner, multimedia', age=43), Row(last_name=u'Vazquez', first_name=u'S
haun', ssn=u'728-08-6560', occupation=u'Health service manager', age=4
1), Row(last_name=u'Hicks', first_name=u'Brian', ssn=u'393-79-9112', oc
cupation=u'Research officer, trade union', age=5), Row(last_name=u'Bail
ey', first_name=u'Tony', ssn=u'787-10-7814', occupation=u'Systems devel
oper', age=19), Row(last_name=u'Martinez', first_name=u'Kristin', ssn=
u'528-75-3096', occupation=u'Materials engineer', age=37), Row(last_nam
e=u'Williams', first_name=u'Calvin', ssn=u'501-65-6512', occupation=u'B
rewing technologist', age=41), Row(last_name=u'Morrison', first_name=
u'Kelly', ssn=u'684-50-3205', occupation=u'Air cabin crew', age=39), Ro
w(last_name=u'Carter', first_name=u'Mary', ssn=u'361-02-6692', occupati
on=u'Emergency planning/management officer', age=0), Row(last_name=u'Jo
nes', first_name=u'Michelle', ssn=u'550-73-1105', occupation=u'Minerals

surveyor', age=27), Row(last_name=u'George', first_name=u'Ashley', ssn=u'855-65-3671', occupation=u'Database administrator', age=42), Row(last_name=u'Brown', first_name=u'Xavier', ssn=u'550-66-3692', occupation=u'Diplomatic Services operational officer', age=40), Row(last_name=u'Blackwell', first_name=u'Darlene', ssn=u'127-46-2963', occupation=u'Arboriculturist', age=44), Row(last_name=u'Ryan', first_name=u'Alexis', ssn=u'587-16-3984', occupation=u'Records manager', age=37), Row(last_name=u'Mueller', first_name=u'Lisa', ssn=u'352-58-1103', occupation=u'Child psychotherapist', age=42), Row(last_name=u'Ross', first_name=u'Thomas', ssn=u'318-58-4031', occupation=u'Production designer, theatre/television/film', age=3), Row(last_name=u'Rosales', first_name=u'Brenda', ssn=u'552-64-1462', occupation=u'Legal secretary', age=44), Row(last_name=u'Watson', first_name=u'Troy', ssn=u'029-16-3218', occupation=u'Information officer', age=17), Row(last_name=u'Brown', first_name=u'Stephen', ssn=u'020-89-8926', occupation=u'Television/film/video producer', age=3), Row(last_name=u'Miller', first_name=u'Brenda', ssn=u'870-78-5179', occupation=u'Scientist, research (physical sciences)', age=46), Row(last_name=u'Blackburn', first_name=u'Elizabeth', ssn=u'114-25-1747', occupation=u'Curator', age=29), Row(last_name=u'Preston', first_name=u'Heidi', ssn=u'749-51-4180', occupation=u'Community development worker', age=16), Row(last_name=u'Taylor', first_name=u'Kim', ssn=u'428-54-8721', occupation=u'Training and development officer', age=27), Row(last_name=u'Taylor', first_name=u'Amy', ssn=u'097-73-4611', occupation=u'Amenity horticulturist', age=21), Row(last_name=u'Klein', first_name=u'Jennifer', ssn=u'311-88-8935', occupation=u'Retail manager', age=27), Row(last_name=u'Figueroa', first_name=u'Barbara', ssn=u'442-80-4571', occupation=u'Theatre manager', age=31), Row(last_name=u'Small', first_name=u'Samantha', ssn=u'659-66-6150', occupation=u'Theme park manager', age=38), Row(last_name=u'Smith', first_name=u'Oscar', ssn=u'043-62-6297', occupation=u'Teacher, secondary school', age=4), Row(last_name=u'Sandoval', first_name=u'Denise', ssn=u'419-19-5311', occupation=u'Holiday representative', age=4), Row(last_name=u'Moore', first_name=u'Diana', ssn=u'202-94-6565', occupation=u'Minerals surveyor', age=6), Row(last_name=u'Johnson', first_name=u'Patricia', ssn=u'059-29-8872', occupation=u'Development worker, community', age=11), Row(last_name=u'Vance', first_name=u'Wesley', ssn=u'116-12-5515', occupation=u'Secondary school teacher', age=14), Row(last_name=u'Pacheco', first_name=u'John', ssn=u'471-96-9503', occupation=u'Risk manager', age=23), Row(last_name=u'Horton', first_name=u'Robert', ssn=u'772-67-2280', occupation=u'Regulatory affairs officer', age=36), Row(last_name=u'Larson', first_name=u'Wendy', ssn=u'395-38-9353', occupation=u'Field seismologist', age=46), Row(last_name=u'Williams', first_name=u'Emily', ssn=u'232-78-4407', occupation=u'Television floor manager', age=14), Row(last_name=u'Contreras', first_name=u'Kevin', ssn=u'430-30-5099', occupation=u'Learning disability nurse', age=5), Row(last_name=u'Brown', first_name=u'Theresa', ssn=u'338-20-4009', occupation=u'Technical brewer', age=35), Row(last_name=u'Hicks', first_name=u'Rebecca', ssn=u'333-41-7417', occupation=u'Conservator, museum/gallery', age=42), Row(last_name=u'Miller', first_name=u'James', ssn=u'899-27-2128', occupation=u'Librarian, academic', age=0), Row(last_name=u'Estrada', first_name=u'Vickie', ssn=u'808-52-5013', occupation=u'Sports coach', age=45), Row(last_name=u'Hart', first_name=u'David', ssn=u'484-49-0392', occupation=u'Communications engineer', age=37), Row(last_name=u'King', first_name=u'Kathryn', ssn=u'752-04-1654', occupation=u'Paramedic', age=28), Row(last_name=u'Durham', first_name=u'John', ssn=u'649-62-7557', occupation=u'Advertising copywriter', age=37), Row(last_name=u'Mahoney', first_name=u'Tami', ssn=u'854-14-9985', occupation=u'Sub', age=39), Row(last_name=u'Jackson', first_name=u'Sandra', ssn

```
=u'441-78-2601', occupation=u'Editor, magazine features', age=46), Row
(last_name=u'Bradley', first_name=u'Jessica', ssn=u'253-56-6348', occup
ation=u'Journalist, broadcasting', age=33), Row(last_name=u'Fowler', fi
rst_name=u'Andrew', ssn=u'402-28-1839', occupation=u'Publishing copy',
 age=17), Row(last_name=u'Moon', first_name=u'Troy', ssn=u'080-51-116
5', occupation=u'Estate manager/land agent', age=6), Row(last_name=u'Cl
ark', first_name=u'Charles', ssn=u'231-44-5618', occupation=u'Marine sc
ientist', age=21), Row(last_name=u'Castro', first_name=u'Christopher',
 ssn=u'727-89-0003', occupation=u'Company secretary', age=23), Row(last
_name=u'Crawford', first_name=u'Clarence', ssn=u'881-91-2884', occupati
on=u'Secondary school teacher', age=35), Row(last_name=u'Evans', first_
name=u'Joseph', ssn=u'716-82-1561', occupation=u'Site engineer', age=
3), Row(last_name=u'Sutton', first_name=u'Melissa', ssn=u'521-05-0402',
 occupation=u'Phytotherapist', age=42), Row(last_name=u'Young', first_n
ame=u'Heidi', ssn=u'146-09-6535', occupation=u'Clinical scientist, hist
ocompatibility and immunogenetics', age=32), Row(last_name=u'Maxwell',
 first_name=u'Amanda', ssn=u'116-45-2790', occupation=u'Designer, inter
ior/spatial', age=0), Row(last_name=u'Romero', first_name=u'Scott', ssn
=u'715-49-4995', occupation=u'Geographical information systems office
r', age=16), Row(last_name=u'Key', first_name=u'Laura', ssn=u'818-61-75
87', occupation=u'Careers information officer', age=27), Row(last_name=
u'Bennett', first_name=u'Joseph', ssn=u'271-50-9273', occupation=u'Comm
unications engineer', age=46), Row(last_name=u'Brady', first_name=u'Sco
tt', ssn=u'603-42-2734', occupation=u'Land/geomatics surveyor', age=4
6), Row(last_name=u'Rivera', first_name=u'Ashley', ssn=u'225-65-2731',
 occupation=u'Designer, graphic', age=9), Row(last_name=u'Williams', fi
rst_name=u'Tanya', ssn=u'313-78-7338', occupation=u'Insurance broker',
 age=4), Row(last_name=u'Taylor', first_name=u'Pamela', ssn=u'671-37-23
91', occupation=u'Theatre manager', age=21), Row(last_name=u'Blevins',
 first_name=u'Stacey', ssn=u'149-16-3465', occupation=u'Operations geol
ogist', age=38), Row(last_name=u'Rogers', first_name=u'Brittney', ssn=
u'321-91-4052', occupation=u'Retail manager', age=13), Row(last_name=
u'Cummings', first_name=u'Valerie', ssn=u'453-38-1363', occupation=u'Ac
countant, chartered certified', age=35), Row(last_name=u'Ryan', first_n
ame=u'Dwayne', ssn=u'056-44-4625', occupation=u'Chartered accountant',
 age=22), Row(last_name=u'Clark', first_name=u'Monique', ssn=u'406-67-4
900', occupation=u'Audiological scientist', age=18), Row(last_name=u'Bo
wman', first_name=u'Kenneth', ssn=u'318-63-5594', occupation=u'Dentis
t', age=22), Row(last_name=u'Johnson', first_name=u'William', ssn=u'630
-14-1134', occupation=u'Runner, broadcasting/film/video', age=45), Row
(last_name=u'Mcdaniel', first_name=u'John', ssn=u'204-52-9964', occupat
ion=u'Garment/textile technologist', age=34), Row(last_name=u'Vazquez',
 first_name=u'Breanna', ssn=u'500-16-4409', occupation=u'Textile design
er', age=43), Row(last_name=u'Hardy', first_name=u'Sierra', ssn=u'065-6
3-9334', occupation=u'Town planner', age=45), Row(last_name=u'Jones', f
irst_name=u'Greg', ssn=u'190-48-4762', occupation=u'Scientist, physiolo
gical', age=18), Row(last_name=u'Clark', first_name=u'Linda', ssn=u'736
-40-7382', occupation=u'Nurse, learning disability', age=10), Row(last_
name=u'George', first_name=u'Emily', ssn=u'077-38-1444', occupation=u'Q
uantity surveyor', age=19), Row(last_name=u'Ponce', first_name=u'Bria
n', ssn=u'520-33-8532', occupation=u'Community arts worker', age=23), R
ow(last_name=u'Reese', first_name=u'Joseph', ssn=u'306-80-5474', occupa
tion=u'Insurance risk surveyor', age=8), Row(last_name=u'Arias', first_
name=u'Kayla', ssn=u'109-90-5374', occupation=u'Theatre stage manager',
 age=12), Row(last_name=u'Davis', first_name=u'Johnathan', ssn=u'545-26
-6597', occupation=u'Brewing technologist', age=12), Row(last_name=u'Go
mez', first_name=u'Lisa', ssn=u'703-62-2800', occupation=u'Intelligence
```

analyst', age=34), Row(last_name=u'Jackson', first_name=u'Jennifer', s
sn=u'749-99-6028', occupation=u'Geophysicist/field seismologist', age=3
8), Row(last_name=u'Malone', first_name=u'Maria', ssn=u'276-81-6111', o
ccupation=u'Personal assistant', age=6), Row(last_name=u'Gonzalez', fir
st_name=u'Dana', ssn=u'694-39-4605', occupation=u'Fashion designer', ag
e=14), Row(last_name=u'Burch', first_name=u'Charles', ssn=u'301-43-072
4', occupation=u'Television production assistant', age=12), Row(last_na
me=u'Cain', first_name=u'Vickie', ssn=u'745-52-1225', occupation=u'Corp
orate treasurer', age=16), Row(last_name=u'Matthews', first_name=u'Jame
s', ssn=u'692-13-9869', occupation=u'Herbalist', age=45), Row(last_name
=u'Diaz', first_name=u'William', ssn=u'142-86-7015', occupation=u'Exhib
itions officer, museum/gallery', age=22), Row(last_name=u'Williams', fi
rst_name=u'Peter', ssn=u'464-30-0294', occupation=u'Animator', age=3),
Row(last_name=u'Miller', first_name=u'Rebecca', ssn=u'366-36-1154', oc
cupation=u'Systems analyst', age=43), Row(last_name=u'Wright', first_na
me=u'James', ssn=u'640-64-8788', occupation=u'Development worker, commu
nity', age=11), Row(last_name=u'Richard', first_name=u'Mr.', ssn=u'028-
53-5208', occupation=u'Buyer, retail', age=13), Row(last_name=u'White',
first_name=u'Grant', ssn=u'331-36-3931', occupation=u'Adult guidance w
orker', age=46), Row(last_name=u'Sims', first_name=u'Christian', ssn=
u'358-52-0437', occupation=u'Designer, multimedia', age=19), Row(last_n
ame=u'Brown', first_name=u'Debra', ssn=u'168-25-4407', occupation=u'Dan
ce movement psychotherapist', age=16), Row(last_name=u'Anderson', first
_name=u'Kathleen', ssn=u'362-51-5389', occupation=u'Tree surgeon', age=
22), Row(last_name=u'Marks', first_name=u'Elizabeth', ssn=u'840-60-656
1', occupation=u'Engineer, mining', age=32), Row(last_name=u'House', fi
rst_name=u'Peter', ssn=u'207-84-2223', occupation=u'Teacher, adult educ
ation', age=36), Row(last_name=u'Farrell', first_name=u'Peter', ssn=u'5
58-43-0487', occupation=u'Statistician', age=1), Row(last_name=u'West',
first_name=u'Julia', ssn=u'842-50-0496', occupation=u'Technical autho
r', age=18), Row(last_name=u'Harris', first_name=u'Allison', ssn=u'453-
23-7614', occupation=u'Therapist, occupational', age=44), Row(last_name
=u'Parker', first_name=u'Amy', ssn=u'605-10-0943', occupation=u'Charter
ed legal executive (England and Wales)', age=2), Row(last_name=u'Canno
n', first_name=u'James', ssn=u'157-33-6280', occupation=u'Psychotherapi
st, child', age=35), Row(last_name=u'Robinson', first_name=u'Richard',
ssn=u'660-95-5623', occupation=u'Surveyor, mining', age=33), Row(last_
name=u'Patterson', first_name=u'Carl', ssn=u'639-62-5566', occupation=
u'Colour technologist', age=38), Row(last_name=u'Bell', first_name=u'Ma
rgaret', ssn=u'314-04-6046', occupation=u'Ship broker', age=27), Row(la
st_name=u'Hudson', first_name=u'Laura', ssn=u'719-47-8562', occupation=
u'Training and development officer', age=43), Row(last_name=u'Mcbride',
first_name=u'Eric', ssn=u'253-01-3546', occupation=u'English as a seco
nd language teacher', age=37), Row(last_name=u'Baker', first_name=u'Sam
uel', ssn=u'006-11-9152', occupation=u'Nutritional therapist', age=10),
Row(last_name=u'Reed', first_name=u'David', ssn=u'581-28-9624', occupa
tion=u'Occupational hygienist', age=10), Row(last_name=u'Nguyen', first
_name=u'Martin', ssn=u'423-88-3932', occupation=u'Manufacturing enginee
r', age=22), Row(last_name=u'Dudley', first_name=u'Katelyn', ssn=u'252-
24-2446', occupation=u'Science writer', age=38), Row(last_name=u'Klin
e', first_name=u'Sarah', ssn=u'485-65-4113', occupation=u'Educational p
sychologist', age=11), Row(last_name=u'Bell', first_name=u'Jessica', ss
n=u'827-84-0528', occupation=u'Recycling officer', age=35), Row(last_na
me=u'Lee', first_name=u'Sandra', ssn=u'551-08-2107', occupation=u'Educa
tion administrator', age=28), Row(last_name=u'Daniel', first_name=u'M
r.', ssn=u'124-62-1947', occupation=u'Interpreter', age=14), Row(last_n
ame=u'Huber', first_name=u'Curtis', ssn=u'330-34-6255', occupation=u'Bu

yer, industrial', age=5), Row(last_name=u'James', first_name=u'Dakota',
 ssn=u'537-96-1324', occupation=u'International aid/development worke
r', age=38), Row(last_name=u'Wilson', first_name=u'Michael', ssn=u'421-
44-5867', occupation=u'Scientific laboratory technician', age=33), Row
(last_name=u'Valencia', first_name=u'Christina', ssn=u'164-38-3408', oc
cupation=u'Chartered management accountant', age=7), Row(last_name=u'Re
id', first_name=u'Shane', ssn=u'236-57-6373', occupation=u'Naval archit
ect', age=23), Row(last_name=u'Morris', first_name=u'Maria', ssn=u'801-
56-8276', occupation=u'Records manager', age=30), Row(last_name=u'Mcfar
land', first_name=u'Diana', ssn=u'790-02-9504', occupation=u'Broadcast
 engineer', age=28), Row(last_name=u'Howard', first_name=u'Eduardo', ss
n=u'817-30-7967', occupation=u'Surveyor, building control', age=37), Ro
w(last_name=u'Perez', first_name=u'Sandra', ssn=u'124-63-7728', occupat
ion=u'Dispensing optician', age=27), Row(last_name=u'Shelton', first_na
me=u'Christine', ssn=u'862-98-9370', occupation=u'Trading standards off
icer', age=14), Row(last_name=u'Johnston', first_name=u'Rachel', ssn=
u'870-03-5211', occupation=u'Bookseller', age=35), Row(last_name=u'Walk
er', first_name=u'Chad', ssn=u'681-97-1216', occupation=u'Meteorologis
t', age=5), Row(last_name=u'Bennett', first_name=u'Eric', ssn=u'826-68-
9239', occupation=u'Teacher, secondary school', age=18), Row(last_name=
u'Robinson', first_name=u'Caitlin', ssn=u'117-23-3103', occupation=u'Ac
tor', age=31), Row(last_name=u'Williams', first_name=u'Kimberly', ssn=
u'124-46-5660', occupation=u'Production engineer', age=0), Row(last_nam
e=u'Weaver', first_name=u'Robin', ssn=u'065-99-4230', occupation=u'Insu
rance underwriter', age=15), Row(last_name=u'Bryant', first_name=u'Chri
stopher', ssn=u'200-15-4213', occupation=u'Graphic designer', age=24),
 Row(last_name=u'Wright', first_name=u'Jacob', ssn=u'278-24-3738', occu
pation=u'Ceramics designer', age=0), Row(last_name=u'Harrison', first_n
ame=u'James', ssn=u'038-21-6567', occupation=u'Public affairs consultan
t', age=43), Row(last_name=u'Ramos', first_name=u'Patrick', ssn=u'268-9
0-6802', occupation=u'Armed forces logistics/support/administrative off
icer', age=39), Row(last_name=u'Flores', first_name=u'Robert', ssn=u'72
5-85-8716', occupation=u'Teacher, secondary school', age=43), Row(last_
name=u'Mayo', first_name=u'Paul', ssn=u'708-93-5822', occupation=u'Manu
facturing engineer', age=28), Row(last_name=u'Parker', first_name=u'Nic
ole', ssn=u'520-67-6094', occupation=u'Occupational hygienist', age=4
0), Row(last_name=u'Rivera', first_name=u'Ryan', ssn=u'697-61-4614', oc
cupation=u'Geographical information systems officer', age=19), Row(last
_name=u'Parker', first_name=u'Donna', ssn=u'030-51-9962', occupation=
u'Make', age=25), Row(last_name=u'Calderon', first_name=u'Jennifer', ss
n=u'346-93-3871', occupation=u'Regulatory affairs officer', age=6), Row
(last_name=u'Harris', first_name=u'Kendra', ssn=u'062-85-8020', occupat
ion=u'Actor', age=8), Row(last_name=u'Lopez', first_name=u'Paula', ssn=
u'379-88-4161', occupation=u'Surveyor, quantity', age=10), Row(last_nam
e=u'Chen', first_name=u'Jason', ssn=u'870-72-4048', occupation=u'Presen
ter, broadcasting', age=27), Row(last_name=u'Mcclure', first_name=u'Kel
ly', ssn=u'098-14-8424', occupation=u'Sports coach', age=11), Row(last_
name=u'Hamilton', first_name=u'Andrea', ssn=u'080-54-4251', occupation=
u'Clinical molecular geneticist', age=31), Row(last_name=u'Werner', fir
st_name=u'Kelsey', ssn=u'874-45-1093', occupation=u'Copy', age=0), Row
(last_name=u'Harris', first_name=u'Angela', ssn=u'743-09-7277', occupat
ion=u'Accommodation manager', age=3), Row(last_name=u'Rogers', first_na
me=u'Thomas', ssn=u'538-86-5046', occupation=u'Production assistant, ra
dio', age=8), Row(last_name=u'Jones', first_name=u'Joel', ssn=u'899-98-
5386', occupation=u'Company secretary', age=37), Row(last_name=u'Howar
d', first_name=u'Jennifer', ssn=u'715-76-3813', occupation=u'Journalis
t, newspaper', age=4), Row(last_name=u'Powers', first_name=u'Christophe

r', ssn=u'138-78-6695', occupation=u'Facilities manager', age=6), Row(last_name=u'Roberson', first_name=u'Jennifer', ssn=u'053-36-0552', occupation=u'Horticultural consultant', age=42), Row(last_name=u'Wells', first_name=u'Jermaine', ssn=u'621-82-8115', occupation=u'Scientist, research (maths)', age=8), Row(last_name=u'Newman', first_name=u'Kevin', ssn=u'201-46-0642', occupation=u'Water engineer', age=12), Row(last_name=u'Mcbride', first_name=u'Dominic', ssn=u'528-35-2678', occupation=u'Surveyor, hydrographic', age=30), Row(last_name=u'Jones', first_name=u'Allison', ssn=u'536-16-6585', occupation=u'Probation officer', age=34), Row(last_name=u'Jones', first_name=u'Larry', ssn=u'178-30-3561', occupation=u'Engineer, electronics', age=32), Row(last_name=u'Smith', first_name=u'Nicole', ssn=u'691-84-6536', occupation=u'Engineer, water', age=23), Row(last_name=u'Johnson', first_name=u'Taylor', ssn=u'453-37-6915', occupation=u'Medical physicist', age=3), Row(last_name=u'Bishop', first_name=u'Jennifer', ssn=u'670-65-9789', occupation=u'Secondary school teacher', age=20), Row(last_name=u'Mullins', first_name=u'Brenda', ssn=u'268-22-0014', occupation=u'Arboriculturist', age=31), Row(last_name=u'Barker', first_name=u'Susan', ssn=u'304-68-2913', occupation=u'Personnel officer', age=37), Row(last_name=u'Bryan', first_name=u'Mary', ssn=u'358-79-5453', occupation=u'Scientist, audiological', age=24), Row(last_name=u'Richardson', first_name=u'Roberta', ssn=u'396-73-0328', occupation=u'Geographical information systems officer', age=41), Row(last_name=u'Yoder', first_name=u'Richard', ssn=u'151-76-2038', occupation=u"Barrister's clerk", age=23), Row(last_name=u'Jones', first_name=u'Jennifer', ssn=u'567-99-4161', occupation=u'Banker', age=0), Row(last_name=u'Webb', first_name=u'Jeffrey', ssn=u'591-78-5683', occupation=u'Designer, jewellery', age=14), Row(last_name=u'Macdonald', first_name=u'William', ssn=u'065-95-4951', occupation=u'Air broker', age=2), Row(last_name=u'Riley', first_name=u'Curtis', ssn=u'783-22-9907', occupation=u'Animator', age=3), Row(last_name=u'Kelly', first_name=u'Lori', ssn=u'583-94-3364', occupation=u'Surveyor, mining', age=8), Row(last_name=u'Cruz', first_name=u'David', ssn=u'213-38-8531', occupation=u'Paediatric nurse', age=9), Row(last_name=u'Brown', first_name=u'Crystal', ssn=u'149-47-1292', occupation=u'Research officer, trade union', age=31), Row(last_name=u'Thompson', first_name=u'Justin', ssn=u'409-03-0918', occupation=u'Theatre manager', age=38), Row(last_name=u'Sanchez', first_name=u'Felicia', ssn=u'674-72-1133', occupation=u'Therapist, horticultural', age=37), Row(last_name=u'Henderson', first_name=u'Emily', ssn=u'848-07-0684', occupation=u'Chartered legal executive (England and Wales)', age=15), Row(last_name=u'Casey', first_name=u'Seth', ssn=u'433-29-3120', occupation=u'Editor, film/video', age=19), Row(last_name=u'Christopher', first_name=u'Mr.', ssn=u'670-90-6147', occupation=u'Magazine features editor', age=37), Row(last_name=u'Johnson', first_name=u'Matthew', ssn=u'676-33-6901', occupation=u'Aid worker', age=14), Row(last_name=u'Hamilton', first_name=u'Michelle', ssn=u'462-75-7918', occupation=u'Dancer', age=17), Row(last_name=u'Johnson', first_name=u'Elizabeth', ssn=u'750-65-2855', occupation=u'Field trials officer', age=45), Row(last_name=u'Chan', first_name=u'William', ssn=u'017-03-8629', occupation=u'Administrator, sports', age=15), Row(last_name=u'Smith', first_name=u'Mercedes', ssn=u'719-65-1615', occupation=u'Naval architect', age=22), Row(last_name=u'Schwartz', first_name=u'Craig', ssn=u'128-60-0010', occupation=u'Telecommunications researcher', age=26), Row(last_name=u'Kent', first_name=u'Roger', ssn=u'755-12-3252', occupation=u'Chartered certified accountant', age=27), Row(last_name=u'Bennett', first_name=u'Sabrina', ssn=u'074-13-2876', occupation=u'Designer, television/film set', age=3), Row(last_name=u'Elliott', first_name=u'Jonathan', ssn=u'455-97-1617', occupation=u'Psychologist, educational', age=16), Row(last_name=u'B

utler', first_name=u'Emily', ssn=u'364-98-0242', occupation=u'Scientis
t, water quality', age=3), Row(last_name=u'Schmidt', first_name=u'Mari
a', ssn=u'760-26-2009', occupation=u'Building surveyor', age=2), Row(la
st_name=u'Rodriguez', first_name=u'William', ssn=u'586-29-3209', occupa
tion=u'Health physicist', age=24), Row(last_name=u'Lucero', first_name=
u'Justin', ssn=u'283-07-3255', occupation=u'Video editor', age=27), Row
(last_name=u'Hughes', first_name=u'Douglas', ssn=u'630-95-4385', occupa
tion=u'Acupuncturist', age=29), Row(last_name=u'Curry', first_name=u'Me
gan', ssn=u'680-82-3197', occupation=u'Designer, furniture', age=34), R
ow(last_name=u'Thomas', first_name=u'Jill', ssn=u'526-74-8887', occupat
ion=u'Patent examiner', age=14), Row(last_name=u'Mueller', first_name=
u'Diana', ssn=u'152-35-5677', occupation=u'Estate manager/land agent',
 age=6), Row(last_name=u'Berg', first_name=u'Sara', ssn=u'297-64-3761',
 occupation=u'Special effects artist', age=24), Row(last_name=u'Davis',
 first_name=u'Daniel', ssn=u'446-28-6886', occupation=u'Corporate treas
urer', age=45), Row(last_name=u'Williams', first_name=u'James', ssn=u'3
93-66-5863', occupation=u'Pharmacist, hospital', age=34), Row(last_name
=u'Mejia', first_name=u'Taylor', ssn=u'804-05-8200', occupation=u'Outdo
or activities/education manager', age=4), Row(last_name=u'Scott', first
_name=u'Lydia', ssn=u'256-51-1313', occupation=u'Sports administrator',
 age=13), Row(last_name=u'Obrien', first_name=u'Victoria', ssn=u'897-67
-5085', occupation=u'Horticultural consultant', age=37), Row(last_name=
u'Tran', first_name=u'Jonathon', ssn=u'505-17-5437', occupation=u'Desig
ner, furniture', age=10), Row(last_name=u'Wagner', first_name=u'Colto
n', ssn=u'761-65-6044', occupation=u'Radiation protection practitione
r', age=9), Row(last_name=u'Sanders', first_name=u'Paula', ssn=u'003-04
-0235', occupation=u'Sound technician, broadcasting/film/video', age=2
9), Row(last_name=u'Walker', first_name=u'Mark', ssn=u'826-18-3580', oc
cupation=u'Personnel officer', age=15), Row(last_name=u'Munoz', first_n
ame=u'Cynthia', ssn=u'263-95-5666', occupation=u'Medical illustrator',
 age=2), Row(last_name=u'Price', first_name=u'Heather', ssn=u'533-16-41
18', occupation=u'Veterinary surgeon', age=33), Row(last_name=u'Rhode
s', first_name=u'Todd', ssn=u'343-89-2083', occupation=u'Insurance risk
 surveyor', age=5), Row(last_name=u'Floyd', first_name=u'Shawn', ssn=
u'710-09-4304', occupation=u'Teacher, primary school', age=42), Row(las
t_name=u'Woods', first_name=u'James', ssn=u'185-58-4962', occupation=
u'Editor, film/video', age=2), Row(last_name=u'Sampson', first_name=u'J
acob', ssn=u'771-19-0106', occupation=u'Clothing/textile technologist',
 age=40), Row(last_name=u'Hernandez', first_name=u'Christian', ssn=u'34
1-97-3400', occupation=u'Public librarian', age=3), Row(last_name=u'Sch
midt', first_name=u'Kristin', ssn=u'395-54-7449', occupation=u'Administ
rator, arts', age=19), Row(last_name=u'Mcmahon', first_name=u'Kevin', s
sn=u'674-95-2613', occupation=u'Chemist, analytical', age=36), Row(last
_name=u'Gray', first_name=u'Lori', ssn=u'157-71-2385', occupation=u'Cat
ering manager', age=37), Row(last_name=u'Cox', first_name=u'Christophe
r', ssn=u'252-31-2615', occupation=u'Editor, commissioning', age=4), Ro
w(last_name=u'Carroll', first_name=u'Michelle', ssn=u'856-29-3633', occ
upation=u'Sports therapist', age=29), Row(last_name=u'Haley', first_nam
e=u'Sheila', ssn=u'475-95-4809', occupation=u'Control and instrumentati
on engineer', age=7), Row(last_name=u'Beasley', first_name=u'Bradley',
 ssn=u'451-47-6421', occupation=u'Quality manager', age=8), Row(last_na
me=u'Baldwin', first_name=u'Aaron', ssn=u'669-88-2716', occupation=u'Im
munologist', age=14), Row(last_name=u'Wise', first_name=u'Annette', ssn
=u'439-93-5298', occupation=u'Dealer', age=4), Row(last_name=u'Sosa', f
irst_name=u'William', ssn=u'222-35-7145', occupation=u'Forest/woodland
 manager', age=43), Row(last_name=u'Rivera', first_name=u'Makayla', ssn
=u'695-85-6924', occupation=u'Careers adviser', age=41), Row(last_name=

u'Baker', first_name=u'Thomas', ssn=u'508-49-0771', occupation=u"Politi
cian's assistant", age=43), Row(last_name=u'Fuentes', first_name=u'Mich
elle', ssn=u'856-83-0904', occupation=u'Geophysical data processor', ag
e=5), Row(last_name=u'Hoffman', first_name=u'Krista', ssn=u'417-54-536
3', occupation=u'Operational investment banker', age=25), Row(last_name
=u'Berg', first_name=u'Brian', ssn=u'403-98-7219', occupation=u'Sales p
romotion account executive', age=12), Row(last_name=u'Martinez', first_
name=u'John', ssn=u'688-26-4944', occupation=u'Biomedical scientist', a
ge=30), Row(last_name=u'Jacobson', first_name=u'William', ssn=u'545-68-
8164', occupation=u'Conference centre manager', age=1), Row(last_name=
u'Carroll', first_name=u'Charles', ssn=u'328-86-5833', occupation=u'Cyt
ogeneticist', age=42), Row(last_name=u'Williams', first_name=u'Erica',
ssn=u'530-63-8796', occupation=u'Surveyor, insurance', age=6), Row(las
t_name=u'Miller', first_name=u'Chelsea', ssn=u'663-07-0197', occupation
=u'Dancer', age=10), Row(last_name=u'Robinson', first_name=u'Alexandr
a', ssn=u'502-81-5673', occupation=u'Garment/textile technologist', age
=36), Row(last_name=u'Williams', first_name=u'Alison', ssn=u'434-45-930
7', occupation=u'Teaching laboratory technician', age=13), Row(last_nam
e=u'Jennifer', first_name=u'Mrs.', ssn=u'708-39-9869', occupation=u'Tax
adviser', age=40), Row(last_name=u'Maxwell', first_name=u'Robert', ssn
=u'645-05-4019', occupation=u'Engineer, aeronautical', age=46), Row(las
t_name=u'Byrd', first_name=u'Jay', ssn=u'090-88-6268', occupation=u'Fin
ancial trader', age=4), Row(last_name=u'Collins', first_name=u'Dean', s
sn=u'592-61-6133', occupation=u'Financial controller', age=6), Row(last
_name=u'Green', first_name=u'Daniel', ssn=u'320-06-9278', occupation=
u'Statistician', age=7), Row(last_name=u'Duncan', first_name=u'Vicki',
ssn=u'033-16-7373', occupation=u'Soil scientist', age=4), Row(last_nam
e=u'Coleman', first_name=u'Leslie', ssn=u'636-38-8824', occupation=u'Te
chnical sales engineer', age=27), Row(last_name=u'Bradford', first_name
=u'Meghan', ssn=u'809-06-4968', occupation=u'Industrial buyer', age=8),
Row(last_name=u'Marshall', first_name=u'Jason', ssn=u'543-63-1842', oc
cupation=u'Nurse, adult', age=13), Row(last_name=u'Martin', first_name=
u'David', ssn=u'706-95-2222', occupation=u'Adult guidance worker', age=
42), Row(last_name=u'Mahoney', first_name=u'Heather', ssn=u'520-67-407
4', occupation=u'Research scientist (medical)', age=33), Row(last_name=
u'Chapman', first_name=u'Autumn', ssn=u'873-19-4037', occupation=u'Muse
um education officer', age=14), Row(last_name=u'Crawford', first_name=
u'Nicholas', ssn=u'488-92-4796', occupation=u'Government social researc
h officer', age=41), Row(last_name=u'Prince', first_name=u'Erin', ssn=
u'358-19-1350', occupation=u'Computer games developer', age=9), Row(las
t_name=u'Davis', first_name=u'Elizabeth', ssn=u'111-98-4444', occupatio
n=u'Conservator, furniture', age=23), Row(last_name=u'Jacobs', first_na
me=u'Amanda', ssn=u'271-05-6491', occupation=u'Training and development
officer', age=15), Row(last_name=u'Diaz', first_name=u'Walter', ssn=
u'301-18-7651', occupation=u'Therapist, music', age=36), Row(last_name=
u'Edwards', first_name=u'Teresa', ssn=u'078-13-7706', occupation=u'Came
ra operator', age=22), Row(last_name=u'Myers', first_name=u'Donna', ssn
=u'844-84-7394', occupation=u'Education officer, community', age=2), Ro
w(last_name=u'Soto', first_name=u'Donald', ssn=u'500-19-5952', occupati
on=u'Librarian, academic', age=35), Row(last_name=u'Jones', first_name=
u'Joseph', ssn=u'573-10-1104', occupation=u'Archaeologist', age=42), Ro
w(last_name=u'Rivera', first_name=u'Melissa', ssn=u'631-32-2105', occup
ation=u'Environmental consultant', age=43), Row(last_name=u'Shah', firs
t_name=u'Debra', ssn=u'432-49-4154', occupation=u'Administrator, charit
ies/voluntary organisations', age=20), Row(last_name=u'Anderson', first
_name=u'Michael', ssn=u'068-42-8597', occupation=u'Engineer, control an
d instrumentation', age=4), Row(last_name=u'Morrison', first_name=u'Ale

xander', ssn=u'533-17-7297', occupation=u'Systems analyst', age=24), Ro
w(last_name=u'Ward', first_name=u'Mark', ssn=u'132-81-1442', occupation
=u'Designer, jewellery', age=14), Row(last_name=u'Ramsey', first_name=
u'David', ssn=u'543-88-2249', occupation=u'Microbiologist', age=35), Ro
w(last_name=u'Moon', first_name=u'Daniel', ssn=u'050-18-7568', occupati
on=u'Fish farm manager', age=38), Row(last_name=u'Anderson', first_name
=u'Johnny', ssn=u'091-66-5255', occupation=u'Clinical embryologist', ag
e=27), Row(last_name=u'Montes', first_name=u'Shannon', ssn=u'122-12-497
1', occupation=u'Education administrator', age=42), Row(last_name=u'Hun
ter', first_name=u'Daniel', ssn=u'019-56-7546', occupation=u'Designer,
industrial/product', age=11), Row(last_name=u'Scott', first_name=u'Emi
ly', ssn=u'682-39-9627', occupation=u'Scientist, forensic', age=43), Ro
w(last_name=u'Gonzalez', first_name=u'Tony', ssn=u'588-47-2104', occupa
tion=u'Trading standards officer', age=14), Row(last_name=u'Hamilton',
first_name=u'Sabrina', ssn=u'585-38-0005', occupation=u'Printmaker', a
ge=16), Row(last_name=u'Booth', first_name=u'Jordan', ssn=u'203-06-446
1', occupation=u'Sports development officer', age=1), Row(last_name=u'R
uiz', first_name=u'Jenna', ssn=u'315-75-1520', occupation=u'Intelligenc
e analyst', age=37), Row(last_name=u'Duffy', first_name=u'Erica', ssn=
u'311-82-1204', occupation=u'Engineer, electrical', age=26), Row(last_n
ame=u'Yang', first_name=u'Jessica', ssn=u'798-03-9741', occupation=u'Ru
ral practice surveyor', age=4), Row(last_name=u'Wright', first_name=u'K
ristina', ssn=u'782-44-3792', occupation=u'Quality manager', age=8), Ro
w(last_name=u'Grant', first_name=u'David', ssn=u'545-95-0881', occupati
on=u'Musician', age=38), Row(last_name=u'Bass', first_name=u'Laura', ss
n=u'119-48-6188', occupation=u'Community education officer', age=19), R
ow(last_name=u'West', first_name=u'Samuel', ssn=u'095-06-9503', occupat
ion=u'Phytotherapist', age=16), Row(last_name=u'Hammond', first_name=
u'James', ssn=u'713-80-9876', occupation=u'Management consultant', age=
30), Row(last_name=u'Zuniga', first_name=u'Matthew', ssn=u'192-93-943
4', occupation=u'Biomedical scientist', age=39), Row(last_name=u'Wilso
n', first_name=u'Mary', ssn=u'577-86-2205', occupation=u'Clinical molec
ular geneticist', age=4), Row(last_name=u'Guerrero', first_name=u'Matth
ew', ssn=u'705-80-9969', occupation=u'Fast food restaurant manager', ag
e=8), Row(last_name=u'Sawyer', first_name=u'Ryan', ssn=u'779-47-6013',
occupation=u'Financial planner', age=46), Row(last_name=u'King', first
_name=u'Stephen', ssn=u'891-27-0290', occupation=u'Television productio
n assistant', age=36), Row(last_name=u'Perry', first_name=u'Jessica', s
sn=u'499-73-0526', occupation=u'Warden/ranger', age=3), Row(last_name=
u'Miller', first_name=u'Jordan', ssn=u'313-81-5712', occupation=u'Resea
rch scientist (maths)', age=26), Row(last_name=u'Berg', first_name=u'Ti
mothy', ssn=u'020-61-3307', occupation=u'Film/video editor', age=41), R
ow(last_name=u'Bryan', first_name=u'Zachary', ssn=u'730-40-6327', occup
ation=u'Lexicographer', age=12), Row(last_name=u'Lopez', first_name=u'K
ristie', ssn=u'867-47-0494', occupation=u'Nurse, mental health', age=1
9), Row(last_name=u'Hoffman', first_name=u'Ricky', ssn=u'219-95-8472',
occupation=u'Clinical embryologist', age=43), Row(last_name=u'Porter',
first_name=u'April', ssn=u'259-47-7815', occupation=u'Aid worker', age
=37), Row(last_name=u'Fuller', first_name=u'Devon', ssn=u'718-86-1093',
occupation=u'Conservator, museum/gallery', age=22), Row(last_name=u'Jo
nes', first_name=u'Lisa', ssn=u'532-15-8799', occupation=u'Armed forces
training and education officer', age=15), Row(last_name=u'Jones', firs
t_name=u'Amy', ssn=u'381-05-5626', occupation=u'Surveyor, insurance', a
ge=46), Row(last_name=u'Hudson', first_name=u'Jody', ssn=u'022-94-214
6', occupation=u'Armed forces technical officer', age=39), Row(last_nam
e=u'Lawson', first_name=u'David', ssn=u'514-63-1557', occupation=u'Desi
gner, graphic', age=5), Row(last_name=u'Rios', first_name=u'Susan', ssn

=u'721-95-4399', occupation=u'Sports administrator', age=34), Row(last_
name=u'Smith', first_name=u'Jennifer', ssn=u'637-75-7111', occupation=
u'IT sales professional', age=3), Row(last_name=u'Jensen', first_name=
u'Louis', ssn=u'380-50-0496', occupation=u'Engineer, electronics', age=
36), Row(last_name=u'Hoover', first_name=u'Alyssa', ssn=u'412-96-3448',
occupation=u'Lawyer', age=32), Row(last_name=u'Hansen', first_name=u'C
hristopher', ssn=u'230-96-3278', occupation=u'Development worker, commu
nity', age=11), Row(last_name=u'Rasmussen', first_name=u'Donald', ssn=
u'719-19-1550', occupation=u'Psychotherapist', age=39), Row(last_name=
u'Perkins', first_name=u'Savannah', ssn=u'591-87-0406', occupation=u'Pu
blic house manager', age=40), Row(last_name=u'Alvarez', first_name=u'Er
ica', ssn=u'644-80-3445', occupation=u'Dispensing optician', age=33), R
ow(last_name=u'King', first_name=u'Amy', ssn=u'513-78-1823', occupation
=u'Television floor manager', age=38), Row(last_name=u'Keith', first_na
me=u'Charles', ssn=u'851-38-6127', occupation=u'Exhibition designer', a
ge=14), Row(last_name=u'Sara', first_name=u'Mrs.', ssn=u'819-74-4557',
occupation=u'Further education lecturer', age=6), Row(last_name=u'Lewi
s', first_name=u'Johnathan', ssn=u'517-74-8755', occupation=u'Careers a
dviser', age=40), Row(last_name=u'Amanda', first_name=u'Mrs.', ssn=u'07
2-46-3343', occupation=u'Furniture conservator/restorer', age=40), Row
(last_name=u'Mitchell', first_name=u'Deborah', ssn=u'637-97-1259', occu
pation=u'Solicitor, Scotland', age=1), Row(last_name=u'Robinson', first
_name=u'Lynn', ssn=u'472-32-7430', occupation=u'Hydrographic surveyor',
age=28), Row(last_name=u'Keller', first_name=u'Melissa', ssn=u'880-31-
3890', occupation=u'Risk manager', age=23), Row(last_name=u'Phillips',
first_name=u'Vanessa', ssn=u'254-63-4515', occupation=u'Film/video edi
tor', age=14), Row(last_name=u'Rose', first_name=u'Eric', ssn=u'475-14-
7331', occupation=u'Tax inspector', age=18), Row(last_name=u'Wolf', fir
st_name=u'David', ssn=u'133-19-3540', occupation=u'Paramedic', age=31),
Row(last_name=u'Michael', first_name=u'Angela', ssn=u'340-13-0379', oc
cupation=u'Clinical embryologist', age=22), Row(last_name=u'Lee', first
_name=u'Debbie', ssn=u'719-37-6361', occupation=u'Psychologist, educati
onal', age=37), Row(last_name=u'Evans', first_name=u'Ivan', ssn=u'790-9
4-4733', occupation=u'Operational investment banker', age=13), Row(last
_name=u'Adams', first_name=u'Philip', ssn=u'708-69-4617', occupation=
u'Primary school teacher', age=1), Row(last_name=u'Wilson', first_name=
u'Jillian', ssn=u'338-07-5774', occupation=u'Designer, industrial/produ
ct', age=26), Row(last_name=u'Brown', first_name=u'Carol', ssn=u'608-81
-8171', occupation=u'Event organiser', age=6), Row(last_name=u'Ryan', f
irst_name=u'Jessica', ssn=u'267-91-5356', occupation=u'Educational psyc
hologist', age=20), Row(last_name=u'Garcia', first_name=u'Valerie', ssn
=u'452-40-8367', occupation=u'Records manager', age=40), Row(last_name=
u'Garcia', first_name=u'Richard', ssn=u'845-81-6471', occupation=u'Grap
hic designer', age=34), Row(last_name=u'Garcia', first_name=u'Maria', s
sn=u'322-67-4299', occupation=u'Lexicographer', age=0), Row(last_name=
u'Garcia', first_name=u'Amy', ssn=u'105-69-1633', occupation=u'Claims i
nspector/assessor', age=17), Row(last_name=u'King', first_name=u'Cind
y', ssn=u'112-47-9366', occupation=u'Personnel officer', age=28), Row(l
ast_name=u'Blevins', first_name=u'Jennifer', ssn=u'064-27-8482', occupa
tion=u'Planning and development surveyor', age=19), Row(last_name=u'You
ng', first_name=u'Tyler', ssn=u'005-71-1940', occupation=u'Public relat
ions account executive', age=19), Row(last_name=u'Dixon', first_name=
u'Catherine', ssn=u'572-85-1397', occupation=u'Artist', age=41), Row(la
st_name=u'Sloan', first_name=u'Lisa', ssn=u'428-19-6998', occupation=
u'Catering manager', age=21), Row(last_name=u'Ball', first_name=u'Eri
c', ssn=u'455-06-0142', occupation=u'Gaffer', age=37), Row(last_name=
u'Barker', first_name=u'Helen', ssn=u'098-08-5414', occupation=u'Advert

ising art director', age=33), Row(last_name=u'Carroll', first_name=u'Pamela', ssn=u'623-11-5649', occupation=u'Geoscientist', age=41), Row(last_name=u'Chandler', first_name=u'Olivia', ssn=u'817-97-8991', occupation=u'Optometrist', age=7), Row(last_name=u'Lee', first_name=u'Heather', ssn=u'673-92-5771', occupation=u'Engineer, maintenance', age=28), Row(last_name=u'Smith', first_name=u'Alyssa', ssn=u'280-59-9235', occupation=u'Environmental education officer', age=40), Row(last_name=u'White', first_name=u'Christopher', ssn=u'803-48-6483', occupation=u'Automotive engineer', age=19), Row(last_name=u'Ross', first_name=u'Eric', ssn=u'551-61-0909', occupation=u'Ergonomist', age=4), Row(last_name=u'Martinez', first_name=u'William', ssn=u'653-89-2459', occupation=u'Patent examiner', age=33), Row(last_name=u'Cohen', first_name=u'Kathryn', ssn=u'770-72-1820', occupation=u'Retail buyer', age=32), Row(last_name=u'Mason', first_name=u'Allison', ssn=u'746-50-0096', occupation=u'Structural engineer', age=36), Row(last_name=u'Romero', first_name=u'Sabrina', ssn=u'243-21-4618', occupation=u'Planning and development surveyor', age=41), Row(last_name=u'Rivera', first_name=u'Katelyn', ssn=u'445-27-3293', occupation=u'Civil Service fast streamer', age=2), Row(last_name=u'Patrick', first_name=u'Patricia', ssn=u'539-03-3784', occupation=u'Scientific laboratory technician', age=10), Row(last_name=u'Nguyen', first_name=u'Jessica', ssn=u'688-99-7471', occupation=u'Restaurant manager', age=33), Row(last_name=u'Meyer', first_name=u'Laurie', ssn=u'202-59-6320', occupation=u'Conservator, museum/gallery', age=42), Row(last_name=u'Morgan', first_name=u'Adrienne', ssn=u'836-39-6345', occupation=u'Social researcher', age=10), Row(last_name=u'Stewart', first_name=u'Robert', ssn=u'654-63-4964', occupation=u'Research scientist (maths)', age=14), Row(last_name=u'Goodwin', first_name=u'Eileen', ssn=u'716-18-0903', occupation=u'Surveyor, commercial/residential', age=32), Row(last_name=u'Mitchell', first_name=u'Jason', ssn=u'637-88-5525', occupation=u'Pilot, airline', age=13), Row(last_name=u'Lucero', first_name=u'Jared', ssn=u'252-62-8603', occupation=u'Technical sales engineer', age=21), Row(last_name=u'Garcia', first_name=u'Stephanie', ssn=u'842-49-3993', occupation=u'Financial planner', age=28), Row(last_name=u'Yates', first_name=u'Phillip', ssn=u'045-59-7097', occupation=u'Engineer, materials', age=12), Row(last_name=u'Moore', first_name=u'Michael', ssn=u'887-20-4847', occupation=u'Lexicographer', age=44), Row(last_name=u'Fleming', first_name=u'Heather', ssn=u'014-71-2033', occupation=u'Physicist, medical', age=1), Row(last_name=u'Torres', first_name=u'Michelle', ssn=u'053-24-7740', occupation=u'Teacher, adult education', age=43), Row(last_name=u'Mullen', first_name=u'Thomas', ssn=u'102-76-8924', occupation=u'Warden/ranger', age=4), Row(last_name=u'Schneider', first_name=u'Zachary', ssn=u'387-01-5703', occupation=u'Secretary/administrator', age=46), Row(last_name=u'Walker', first_name=u'Hannah', ssn=u'824-48-9556', occupation=u'Drilling engineer', age=43), Row(last_name=u'Edwards', first_name=u'Mary', ssn=u'868-69-7049', occupation=u'Control and instrumentation engineer', age=26), Row(last_name=u'Davis', first_name=u'Katherine', ssn=u'787-77-8216', occupation=u'Therapeutic radiographer', age=0), Row(last_name=u'Lindsey', first_name=u'Joshua', ssn=u'724-24-2317', occupation=u'Pensions consultant', age=6), Row(last_name=u'Barry', first_name=u'Brenda', ssn=u'023-22-4055', occupation=u'Volunteer coordinator', age=7), Row(last_name=u'Mays', first_name=u'Michael', ssn=u'468-70-9960', occupation=u'Editorial assistant', age=41), Row(last_name=u'Newman', first_name=u'Elizabeth', ssn=u'664-60-2962', occupation=u'Trade mark attorney', age=39), Row(last_name=u'Beasley', first_name=u'Collin', ssn=u'195-58-6281', occupation=u'Secretary/administrator', age=30), Row(last_name=u'Cooley', first_name=u'Jermaine', ssn=u'639-07-2123', occupation=u'Production designer, theatre/television/film', age=31), Row(la

st_name=u'Smith', first_name=u'Holly', ssn=u'831-58-6924', occupation=
u'Conservation officer, nature', age=6), Row(last_name=u'Bradley', firs
t_name=u'Lisa', ssn=u'767-34-0095', occupation=u'Educational psychologi
st', age=33), Row(last_name=u'Singleton', first_name=u'Jennifer', ssn=
u'652-23-8827', occupation=u'Tax inspector', age=2), Row(last_name=u'Ho
ward', first_name=u'Kyle', ssn=u'627-94-5673', occupation=u'Legal execu
tive', age=45), Row(last_name=u'Ball', first_name=u'Andrew', ssn=u'464-
68-1938', occupation=u'Geophysicist/field seismologist', age=25), Row(l
ast_name=u'Barker', first_name=u'Edward', ssn=u'041-83-0268', occupatio
n=u'Health promotion specialist', age=28), Row(last_name=u'Nash', first
_name=u'Monica', ssn=u'111-45-5641', occupation=u'Health service manage
r', age=23), Row(last_name=u'Mosley', first_name=u'Sandra', ssn=u'638-6
3-4130', occupation=u'Emergency planning/management officer', age=1), R
ow(last_name=u'Marshall', first_name=u'Sheena', ssn=u'148-98-1438', occ
upation=u'Quantity surveyor', age=6), Row(last_name=u'Brown', first_nam
e=u'Jacqueline', ssn=u'152-52-9084', occupation=u'Engineer, control and
 instrumentation', age=28), Row(last_name=u'Ellis', first_name=u'Dawn',
 ssn=u'526-86-2605', occupation=u'Geophysical data processor', age=31),
 Row(last_name=u'Garcia', first_name=u'Kathy', ssn=u'596-99-7952', occu
pation=u'Air traffic controller', age=45), Row(last_name=u'Matthews', f
irst_name=u'John', ssn=u'195-76-1883', occupation=u'Optician, dispensin
g', age=9), Row(last_name=u'Dennis', first_name=u'Sean', ssn=u'270-27-6
714', occupation=u'Tour manager', age=4), Row(last_name=u'John', first_
name=u'Dr.', ssn=u'842-76-0029', occupation=u'Estate manager/land agen
t', age=13), Row(last_name=u'Reese', first_name=u'Sara', ssn=u'516-40-7
996', occupation=u'Engineer, water', age=16), Row(last_name=u'Mcdonal
d', first_name=u'Garrett', ssn=u'724-01-1415', occupation=u'Translato
r', age=26), Row(last_name=u'Chandler', first_name=u'Scott', ssn=u'114-
63-9209', occupation=u'Community development worker', age=23), Row(last
_name=u'Harris', first_name=u'Michael', ssn=u'118-33-6781', occupation=
u'Public librarian', age=23), Row(last_name=u'Warren', first_name=u'Sab
rina', ssn=u'562-63-5065', occupation=u'Geoscientist', age=30), Row(las
t_name=u'Larson', first_name=u'Susan', ssn=u'296-04-2261', occupation=
u'Games developer', age=0), Row(last_name=u'Frazier', first_name=u'Eri
n', ssn=u'573-28-7017', occupation=u'Public relations account executiv
e', age=25), Row(last_name=u'Scott', first_name=u'Jane', ssn=u'015-36-3
637', occupation=u'Logistics and distribution manager', age=41), Row(la
st_name=u'Thornton', first_name=u'Sabrina', ssn=u'520-52-5562', occupat
ion=u'Statistician', age=37), Row(last_name=u'Bass', first_name=u'Joh
n', ssn=u'463-96-8977', occupation=u'Estate agent', age=3), Row(last_na
me=u'Mendoza', first_name=u'Angela', ssn=u'236-04-9788', occupation=u'G
ames developer', age=12), Row(last_name=u'Smith', first_name=u'Natash
a', ssn=u'204-14-9422', occupation=u'Youth worker', age=17), Row(last_n
ame=u'Maxwell', first_name=u'Teresa', ssn=u'074-38-9480', occupation=
u'Sport and exercise psychologist', age=25), Row(last_name=u'Waters', f
irst_name=u'Donald', ssn=u'096-45-3721', occupation=u'Environmental man
ager', age=39), Row(last_name=u'Young', first_name=u'Briana', ssn=u'150
-65-4031', occupation=u'Research scientist (medical)', age=6), Row(last
_name=u'Jennings', first_name=u'Susan', ssn=u'244-56-1851', occupation=
u'Special educational needs teacher', age=23), Row(last_name=u'Tucker',
 first_name=u'Ryan', ssn=u'414-20-2526', occupation=u'Solicitor, Scotla
nd', age=20), Row(last_name=u'Beck', first_name=u'Selena', ssn=u'398-67
-2983', occupation=u'Dancer', age=6), Row(last_name=u'Sanchez', first_n
ame=u'Robert', ssn=u'564-83-5239', occupation=u'Investment banker, oper
ational', age=20), Row(last_name=u'Bryant', first_name=u'Amy', ssn=u'63
7-41-8181', occupation=u'Veterinary surgeon', age=30), Row(last_name=
u'Ball', first_name=u'Nicole', ssn=u'860-38-0615', occupation=u'Phytoth

erapist', age=38), Row(last_name=u'Ford', first_name=u'James', ssn=u'85
1-84-4299', occupation=u'Firefighter', age=1), Row(last_name=u'David',
 first_name=u'Dr.', ssn=u'050-65-9131', occupation=u'Microbiologist', a
ge=31), Row(last_name=u'Yang', first_name=u'Amy', ssn=u'777-56-3701', o
ccupation=u'Broadcast presenter', age=39), Row(last_name=u'Burgess', fi
rst_name=u'Kimberly', ssn=u'387-56-1589', occupation=u'Materials engine
er', age=4), Row(last_name=u'Christensen', first_name=u'Mary', ssn=u'69
9-35-7852', occupation=u'Museum/gallery exhibitions officer', age=26),
 Row(last_name=u'Jones', first_name=u'Paul', ssn=u'387-01-9488', occupa
tion=u'Marketing executive', age=5), Row(last_name=u'Harmon', first_nam
e=u'Omar', ssn=u'178-14-1222', occupation=u'Teacher, special educationa
l needs', age=44), Row(last_name=u'Farrell', first_name=u'Angela', ssn=
u'849-11-9970', occupation=u'Radio producer', age=26), Row(last_name=
u'Riggs', first_name=u'Jennifer', ssn=u'200-73-9468', occupation=u'Pres
s sub', age=37), Row(last_name=u'Smith', first_name=u'Lisa', ssn=u'347-
29-5583', occupation=u'Insurance claims handler', age=20), Row(last_nam
e=u'Ward', first_name=u'Cory', ssn=u'348-06-5863', occupation=u'Materia
ls engineer', age=1), Row(last_name=u'Crawford', first_name=u'Amy', ssn
=u'438-37-7868', occupation=u'Psychologist, clinical', age=35), Row(las
t_name=u'Chavez', first_name=u'Cynthia', ssn=u'892-43-7127', occupation
=u'Designer, multimedia', age=35), Row(last_name=u'Burke', first_name=
u'Tina', ssn=u'170-79-5622', occupation=u'Engineer, broadcasting (opera
tions)', age=36), Row(last_name=u'Gibson', first_name=u'Patrick', ssn=
u'591-67-0168', occupation=u'Visual merchandiser', age=32), Row(last_na
me=u'Park', first_name=u'Craig', ssn=u'639-41-5435', occupation=u'Geoph
ysicist/field seismologist', age=16), Row(last_name=u'Sandoval', first_
name=u'Justin', ssn=u'332-44-9275', occupation=u'Best boy', age=31), Ro
w(last_name=u'Hawkins', first_name=u'Elizabeth', ssn=u'674-98-7986', oc
cupation=u'Academic librarian', age=6), Row(last_name=u'Miller', first_
name=u'William', ssn=u'561-37-9150', occupation=u'Field trials office
r', age=32), Row(last_name=u'Miller', first_name=u'Amy', ssn=u'415-37-2
520', occupation=u'Engineer, electronics', age=29), Row(last_name=u'Bar
nes', first_name=u'Lawrence', ssn=u'437-89-6430', occupation=u'Multimed
ia specialist', age=30), Row(last_name=u'Bell', first_name=u'Denise', s
sn=u'212-39-5931', occupation=u'Programmer, applications', age=26), Row
(last_name=u'Scott', first_name=u'Shelia', ssn=u'642-40-0274', occupati
on=u'Energy manager', age=37), Row(last_name=u'Gordon', first_name=u'St
ephanie', ssn=u'583-13-0708', occupation=u'Tax adviser', age=8), Row(la
st_name=u'Bates', first_name=u'Joshua', ssn=u'868-97-5288', occupation=
u'Accounting technician', age=1), Row(last_name=u'Simmons', first_name=
u'Tyler', ssn=u'247-25-5884', occupation=u'Television/film/video produc
er', age=29), Row(last_name=u'Sullivan', first_name=u'Robert', ssn=u'12
3-95-9469', occupation=u'Surveyor, minerals', age=3), Row(last_name=u'H
ill', first_name=u'Dale', ssn=u'853-25-2710', occupation=u'Librarian, p
ublic', age=39), Row(last_name=u'Solis', first_name=u'Danielle', ssn=
u'126-64-8892', occupation=u'Medical laboratory scientific officer', ag
e=18), Row(last_name=u'Mendoza', first_name=u'Amber', ssn=u'414-85-236
7', occupation=u'Music tutor', age=26), Row(last_name=u'Peterson', firs
t_name=u'Kevin', ssn=u'362-14-9869', occupation=u'Orthoptist', age=25),
 Row(last_name=u'Zavala', first_name=u'Mark', ssn=u'711-78-6464', occup
ation=u'Lobbyist', age=2), Row(last_name=u'Jones', first_name=u'Natha
n', ssn=u'532-32-4405', occupation=u'Forensic scientist', age=6), Row(l
ast_name=u'Moore', first_name=u'Jillian', ssn=u'432-67-9845', occupatio
n=u'Stage manager', age=35), Row(last_name=u'Hopkins', first_name=u'Kev
in', ssn=u'875-71-0175', occupation=u'Programmer, systems', age=40), Ro
w(last_name=u'Davis', first_name=u'Jason', ssn=u'549-66-5506', occupati
on=u'Therapeutic radiographer', age=6), Row(last_name=u'Fisher', first_

name=u'Tina', ssn=u'869-66-1868', occupation=u'Investment banker, corpo
rate', age=35), Row(last_name=u'Fowler', first_name=u'Stephanie', ssn=
u'286-17-0346', occupation=u'Purchasing manager', age=28), Row(last_nam
e=u'Sims', first_name=u'Bryan', ssn=u'768-18-0903', occupation=u'Applic
ations developer', age=14), Row(last_name=u'Berry', first_name=u'Meaga
n', ssn=u'235-85-9209', occupation=u'Toxicologist', age=34), Row(last_n
ame=u'Perez', first_name=u'Jennifer', ssn=u'108-63-6279', occupation=
u'Government social research officer', age=44), Row(last_name=u'Gray',
first_name=u'Stephanie', ssn=u'273-11-8534', occupation=u'Engineer, bu
ilding services', age=21), Row(last_name=u'Brown', first_name=u'Cynthi
a', ssn=u'643-78-4736', occupation=u'Administrator, local government',
age=6), Row(last_name=u'King', first_name=u'Christina', ssn=u'007-72-3
608', occupation=u'Therapist, art', age=21), Row(last_name=u'Hill', fir
st_name=u'Lisa', ssn=u'832-54-4960', occupation=u'Development worker, c
ommunity', age=32), Row(last_name=u'Watts', first_name=u'Amanda', ssn=
u'702-98-7106', occupation=u'Podiatrist', age=0), Row(last_name=u'Mcdow
ell', first_name=u'John', ssn=u'598-44-0861', occupation=u'Human resour
ces officer', age=8), Row(last_name=u'Mullins', first_name=u'Casey', ss
n=u'415-43-7594', occupation=u'Solicitor', age=15), Row(last_name=u'Wil
son', first_name=u'Larry', ssn=u'217-61-0637', occupation=u'Therapist,
art', age=45), Row(last_name=u'Phelps', first_name=u'Thomas', ssn=u'65
4-40-3738', occupation=u'Financial controller', age=43), Row(last_name=
u'Stone', first_name=u'Erin', ssn=u'484-70-4761', occupation=u'Scientif
ic laboratory technician', age=29), Row(last_name=u'Ross', first_name=
u'Christina', ssn=u'144-28-6087', occupation=u'Clinical molecular genet
icist', age=28), Row(last_name=u'Hansen', first_name=u'Andrew', ssn=u'0
58-83-7102', occupation=u'Industrial/product designer', age=40), Row(la
st_name=u'Hogan', first_name=u'Joseph', ssn=u'363-09-9705', occupation=
u'Scientist, research (maths)', age=35), Row(last_name=u'Jones', first_
name=u'Andrew', ssn=u'890-43-8664', occupation=u'Health and safety advi
ser', age=24), Row(last_name=u'Shaw', first_name=u'Ashley', ssn=u'817-7
9-0940', occupation=u'Land', age=4), Row(last_name=u'Cannon', first_nam
e=u'Justin', ssn=u'888-32-7305', occupation=u'Counsellor', age=35), Row
(last_name=u'Robbins', first_name=u'Marisa', ssn=u'174-92-7323', occupa
tion=u'Chiropodist', age=3), Row(last_name=u'Flowers', first_name=u'Joh
n', ssn=u'148-12-7993', occupation=u'Administrator, local government',
age=27), Row(last_name=u'Cameron', first_name=u'Nicole', ssn=u'715-42-
1789', occupation=u'Financial adviser', age=22), Row(last_name=u'Bank
s', first_name=u'Monica', ssn=u'331-11-6797', occupation=u'Toxicologis
t', age=4), Row(last_name=u'Lawson', first_name=u'Joseph', ssn=u'555-31
-1481', occupation=u'Garment/textile technologist', age=33), Row(last_n
ame=u'Martinez', first_name=u'Heather', ssn=u'740-74-0857', occupation=
u'Electronics engineer', age=8), Row(last_name=u'Barber', first_name=
u'Margaret', ssn=u'154-30-0696', occupation=u'Rural practice surveyor',
age=28), Row(last_name=u'Miller', first_name=u'Norma', ssn=u'024-93-08
96', occupation=u'Multimedia programmer', age=31), Row(last_name=u'Hale
y', first_name=u'Michael', ssn=u'697-98-1434', occupation=u'Video edito
r', age=34), Row(last_name=u'Thomas', first_name=u'Joseph', ssn=u'510-0
1-3893', occupation=u'Horticulturist, amenity', age=45), Row(last_name=
u'Jackson', first_name=u'Michael', ssn=u'632-88-6245', occupation=u'Cab
in crew', age=36), Row(last_name=u'Moreno', first_name=u'Jennifer', ssn
=u'650-91-5649', occupation=u'Engineer, maintenance', age=26), Row(last
_name=u'Nash', first_name=u'Victor', ssn=u'254-32-7338', occupation=u'C
opywriter, advertising', age=38), Row(last_name=u'Potts', first_name=
u'Patricia', ssn=u'652-09-8709', occupation=u'Control and instrumentati
on engineer', age=22), Row(last_name=u'Hernandez', first_name=u'James',
ssn=u'042-31-1139', occupation=u'Accounting technician', age=5), Row(l

ast_name=u'Williams', first_name=u'Tiffany', ssn=u'205-52-4417', occupa
tion=u'Occupational therapist', age=19), Row(last_name=u'Joseph', first
_name=u'Mr.', ssn=u'131-72-6516', occupation=u'Veterinary surgeon', age
=40), Row(last_name=u'Smith', first_name=u'Todd', ssn=u'151-11-4338', o
ccupation=u'Social worker', age=42), Row(last_name=u'Avery', first_name
=u'Shaun', ssn=u'018-45-9101', occupation=u'Local government officer',
 age=42), Row(last_name=u'Rodgers', first_name=u'Melissa', ssn=u'594-14
-5336', occupation=u'Land', age=19), Row(last_name=u'Pittman', first_na
me=u'Calvin', ssn=u'639-41-1506', occupation=u'Armed forces technical o
fficer', age=28), Row(last_name=u'Lara', first_name=u'Jennifer', ssn=
u'570-84-7839', occupation=u'Therapist, drama', age=35), Row(last_name=
u'Love', first_name=u'Todd', ssn=u'288-08-4327', occupation=u'Surveyor,
 insurance', age=46), Row(last_name=u'Ortiz', first_name=u'Mary', ssn=
u'112-73-8850', occupation=u'Therapist, art', age=2), Row(last_name=u'A
dams', first_name=u'Nancy', ssn=u'039-62-6784', occupation=u'Copy', age
=21), Row(last_name=u'Logan', first_name=u'Mark', ssn=u'186-14-1611', o
ccupation=u'Surveyor, commercial/residential', age=1), Row(last_name=
u'Munoz', first_name=u'Laura', ssn=u'776-13-4126', occupation=u'Social
 research officer, government', age=2), Row(last_name=u'Chandler', firs
t_name=u'Thomas', ssn=u'121-93-9410', occupation=u'Estate agent', age=1
6), Row(last_name=u'Reynolds', first_name=u'Brian', ssn=u'779-96-1263',
 occupation=u'Radiation protection practitioner', age=40), Row(last_nam
e=u'Reeves', first_name=u'Deborah', ssn=u'817-61-4228', occupation=u'Le
xicographer', age=0), Row(last_name=u'Tammy', first_name=u'Dr.', ssn=
u'894-96-7532', occupation=u'Personnel officer', age=10), Row(last_name
=u'Zhang', first_name=u'Tiffany', ssn=u'021-84-7832', occupation=u'Medi
a planner', age=8), Row(last_name=u'Short', first_name=u'Kerry', ssn=
u'404-16-9001', occupation=u'Immigration officer', age=31), Row(last_na
me=u'Miller', first_name=u'Michael', ssn=u'338-80-3491', occupation=u'B
uilding surveyor', age=1), Row(last_name=u'Vazquez', first_name=u'Eri
c', ssn=u'295-52-0903', occupation=u'Catering manager', age=4), Row(las
t_name=u'White', first_name=u'Jacob', ssn=u'797-69-0117', occupation=
u'Set designer', age=7), Row(last_name=u'Simpson', first_name=u'Sarah',
 ssn=u'785-36-2649', occupation=u'Telecommunications researcher', age=4
5), Row(last_name=u'Smith', first_name=u'Kristi', ssn=u'576-95-0860', o
ccupation=u'Chartered loss adjuster', age=46), Row(last_name=u'Martin',
 first_name=u'Michael', ssn=u'703-05-0906', occupation=u'Copywriter, ad
vertising', age=27), Row(last_name=u'Murray', first_name=u'Kayla', ssn=
u'597-01-8445', occupation=u'Psychologist, occupational', age=25), Row
(last_name=u'Garcia', first_name=u'John', ssn=u'124-49-8510', occupatio
n=u'Sales professional, IT', age=31), Row(last_name=u'Torres', first_na
me=u'Michael', ssn=u'706-32-2774', occupation=u'Buyer, industrial', age
=26), Row(last_name=u'Johnson', first_name=u'Scott', ssn=u'236-92-398
0', occupation=u'Programme researcher, broadcasting/film/video', age=1
0), Row(last_name=u'Dawson', first_name=u'Wayne', ssn=u'538-09-7503', o
ccupation=u'Scientist, clinical (histocompatibility and immunogenetic
s)', age=37), Row(last_name=u'Palmer', first_name=u'Charles', ssn=u'423
-13-5304', occupation=u'Landscape architect', age=24), Row(last_name=
u'Jessica', first_name=u'Mrs.', ssn=u'363-27-1536', occupation=u"Barris
ter's clerk", age=43), Row(last_name=u'Jones', first_name=u'Alexander',
 ssn=u'342-99-6075', occupation=u'Fashion designer', age=42), Row(last_
name=u'Maxwell', first_name=u'Laura', ssn=u'411-31-9372', occupation=
u'Housing manager/officer', age=11), Row(last_name=u'Guzman', first_nam
e=u'Suzanne', ssn=u'548-60-8645', occupation=u'Operational researcher',
 age=19), Row(last_name=u'Caldwell', first_name=u'John', ssn=u'595-37-9
993', occupation=u'Furniture designer', age=26), Row(last_name=u'Gibso
n', first_name=u'David', ssn=u'154-69-3863', occupation=u'Television/fi

lm/video producer', age=17), Row(last_name=u'Davis', first_name=u'Edwar
d', ssn=u'541-40-2130', occupation=u'Clinical psychologist', age=10), R
ow(last_name=u'Gamble', first_name=u'Jack', ssn=u'098-12-6250', occupat
ion=u'Investment banker, corporate', age=9), Row(last_name=u'Lawson', f
irst_name=u'Audrey', ssn=u'236-80-2241', occupation=u'Airline pilot', a
ge=42), Row(last_name=u'Mendez', first_name=u'Beth', ssn=u'170-17-807
5', occupation=u'Field trials officer', age=31), Row(last_name=u'Roger
s', first_name=u'Joshua', ssn=u'180-95-6612', occupation=u'Administrato
r, Civil Service', age=2), Row(last_name=u'Tran', first_name=u'Stephani
e', ssn=u'049-93-5421', occupation=u'Energy engineer', age=38), Row(las
t_name=u'Miller', first_name=u'Stephanie', ssn=u'821-63-9253', occupati
on=u'Diplomatic Services operational officer', age=33), Row(last_name=
u'Dawson', first_name=u'Jessica', ssn=u'831-75-5290', occupation=u'Publ
ishing rights manager', age=43), Row(last_name=u'Stone', first_name=u'C
hristina', ssn=u'640-82-6541', occupation=u'Lecturer, further educatio
n', age=6), Row(last_name=u'Oliver', first_name=u'Jesus', ssn=u'213-10-
8080', occupation=u'Publishing copy', age=45), Row(last_name=u'Erickso
n', first_name=u'David', ssn=u'480-42-7266', occupation=u'Planning and
 development surveyor', age=2), Row(last_name=u'Johnson', first_name=
u'Gina', ssn=u'714-22-0824', occupation=u'Civil Service fast streamer',
 age=25), Row(last_name=u'Skinner', first_name=u'David', ssn=u'122-92-9
097', occupation=u'Community arts worker', age=24), Row(last_name=u'Gut
ierrez', first_name=u'Paula', ssn=u'521-19-4741', occupation=u'Enginee
r, structural', age=20), Row(last_name=u'Smith', first_name=u'William',
 ssn=u'718-58-7856', occupation=u'Planning and development surveyor', a
ge=15), Row(last_name=u'Coffey', first_name=u'Kimberly', ssn=u'044-35-8
014', occupation=u'Chemical engineer', age=22), Row(last_name=u'Todd',
 first_name=u'Troy', ssn=u'664-25-2579', occupation=u'Planning and deve
lopment surveyor', age=31), Row(last_name=u'Morgan', first_name=u'Krist
a', ssn=u'208-14-3372', occupation=u'Careers adviser', age=4), Row(last
_name=u'Turner', first_name=u'Shane', ssn=u'167-51-5385', occupation=
u'Chartered loss adjuster', age=19), Row(last_name=u'Wilcox', first_nam
e=u'David', ssn=u'168-94-8331', occupation=u'Chartered management accou
ntant', age=45), Row(last_name=u'Mcdonald', first_name=u'Kara', ssn=u'3
29-45-3753', occupation=u'Pension scheme manager', age=17), Row(last_na
me=u'Bryan', first_name=u'Jeffrey', ssn=u'362-90-5067', occupation=u'Ge
neticist, molecular', age=1), Row(last_name=u'Ross', first_name=u'Richa
rd', ssn=u'521-17-9728', occupation=u'Programmer, multimedia', age=35),
 Row(last_name=u'Martin', first_name=u'Troy', ssn=u'843-87-4270', occup
ation=u'Lecturer, higher education', age=27), Row(last_name=u'Burnett',
 first_name=u'Linda', ssn=u'324-81-5416', occupation=u'Librarian, publi
c', age=2), Row(last_name=u'Watson', first_name=u'Keith', ssn=u'298-05-
2374', occupation=u'Radiographer, diagnostic', age=35), Row(last_name=
u'Wilson', first_name=u'David', ssn=u'851-46-0841', occupation=u'Databa
se administrator', age=13), Row(last_name=u'Nunez', first_name=u'Scot
t', ssn=u'854-59-7667', occupation=u'Administrator, arts', age=31), Row
(last_name=u'Huff', first_name=u'Ashley', ssn=u'670-16-3077', occupatio
n=u'Catering manager', age=17), Row(last_name=u'Rogers', first_name=u'K
enneth', ssn=u'259-71-4330', occupation=u'Designer, interior/spatial',
 age=30), Row(last_name=u'Perez', first_name=u'Melissa', ssn=u'068-85-8
894', occupation=u'Podiatrist', age=20), Row(last_name=u'Gibson', first
_name=u'Luis', ssn=u'194-21-2853', occupation=u'Web designer', age=28),
 Row(last_name=u'Miller', first_name=u'James', ssn=u'857-48-2470', occu
pation=u'Teacher, music', age=6), Row(last_name=u'Mueller', first_name=
u'Pamela', ssn=u'090-60-6129', occupation=u'Environmental manager', age
=34), Row(last_name=u'Rodgers', first_name=u'Jasmine', ssn=u'257-39-017
3', occupation=u'Investment banker, corporate', age=11), Row(last_name=

```
u'Graham', first_name=u'Robert', ssn=u'507-45-4466', occupation=u'Embry
ologist, clinical', age=19), Row(last_name=u'Martin', first_name=u'Dian
a', ssn=u'093-86-9798', occupation=u'Secondary school teacher', age=2
5), Row(last_name=u'Cabrera', first_name=u'Tristan', ssn=u'421-46-246
8', occupation=u'Sports administrator', age=0), Row(last_name=u'Dickso
n', first_name=u'Heather', ssn=u'611-23-8517', occupation=u'Scientist,
 water quality', age=38), Row(last_name=u'Mckinney', first_name=u'Melis
sa', ssn=u'101-39-1432', occupation=u'Regulatory affairs officer', age=
13), Row(last_name=u'Brewer', first_name=u'Dennis', ssn=u'186-87-0304',
 occupation=u'Speech and language therapist', age=10), Row(last_name=
u'Kirk', first_name=u'Timothy', ssn=u'168-07-0940', occupation=u'Geophy
sicist/field seismologist', age=2), Row(last_name=u'Bass', first_name=
u'Tracy', ssn=u'306-19-8921', occupation=u'Illustrator', age=42), Row(l
ast_name=u'Simmons', first_name=u'Joshua', ssn=u'753-67-1241', occupati
on=u'Advice worker', age=37), Row(last_name=u'Richard', first_name=u'M
r.', ssn=u'130-22-9157', occupation=u'Engineer, chemical', age=3), Row
(last_name=u'Simon', first_name=u'Ashley', ssn=u'512-48-7412', occupati
on=u'Lighting technician, broadcasting/film/video', age=20), Row(last_n
ame=u'Chen', first_name=u'Stacy', ssn=u'199-37-6266', occupation=u'Wast
e management officer', age=15), Row(last_name=u'Aguilar', first_name=
u'Christopher', ssn=u'303-64-4132', occupation=u'Ambulance person', age
=6), Row(last_name=u'Bennett', first_name=u'Renee', ssn=u'239-18-4137',
 occupation=u'Producer, televisio
```