**AI Based Diabetes Prediction Model**

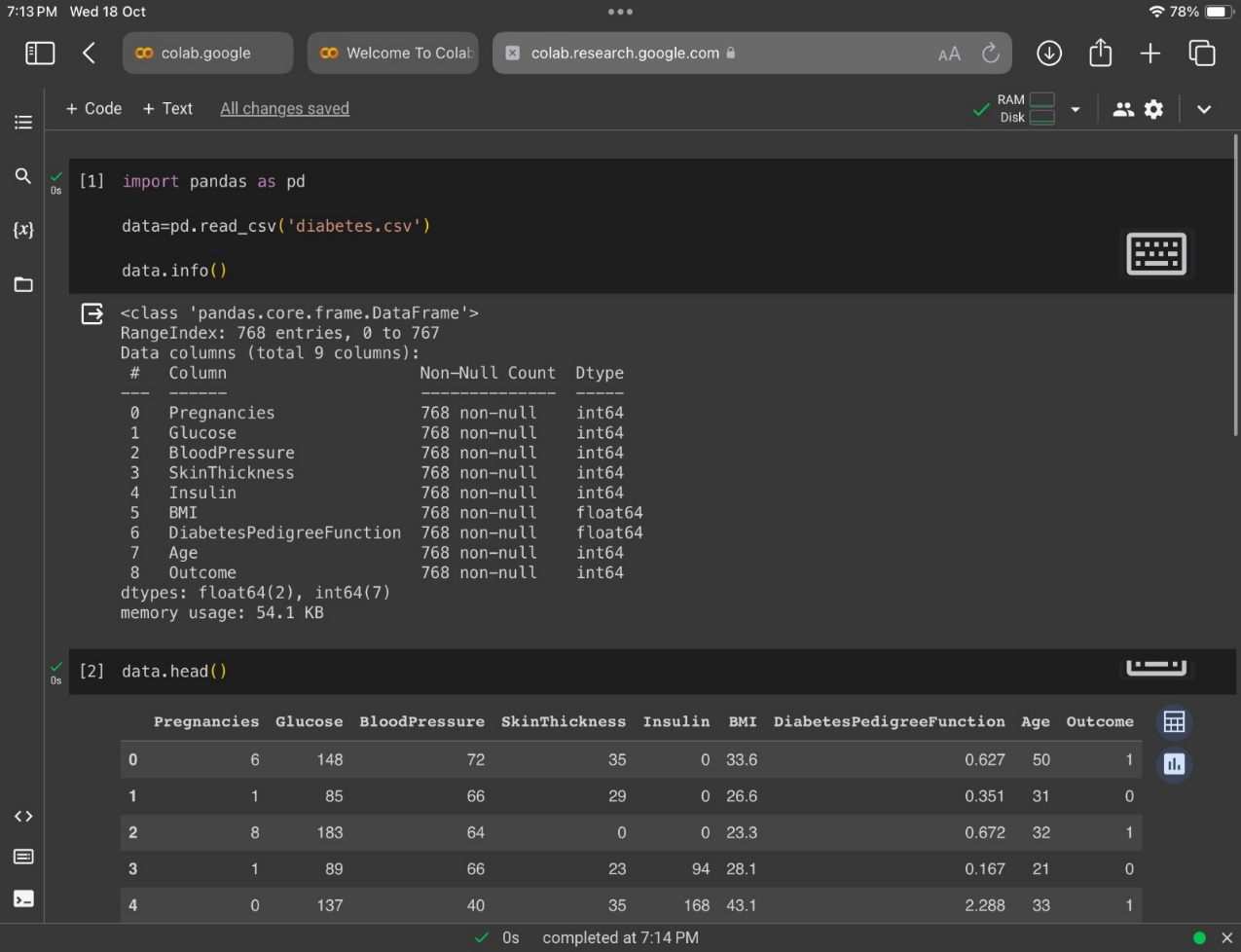
**Phase-3: Development-1**

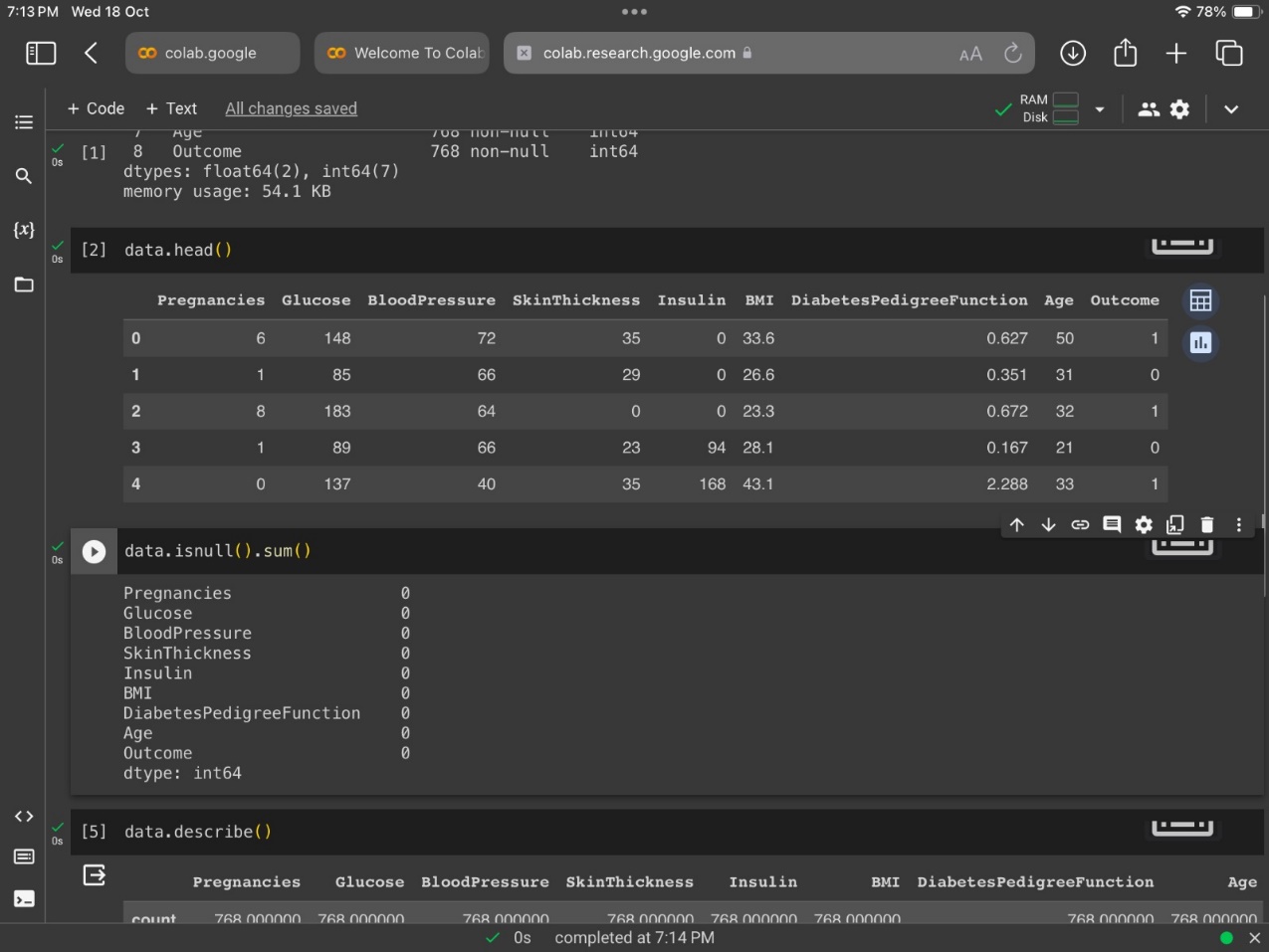
1. **Importing the Dataset and describing it.**

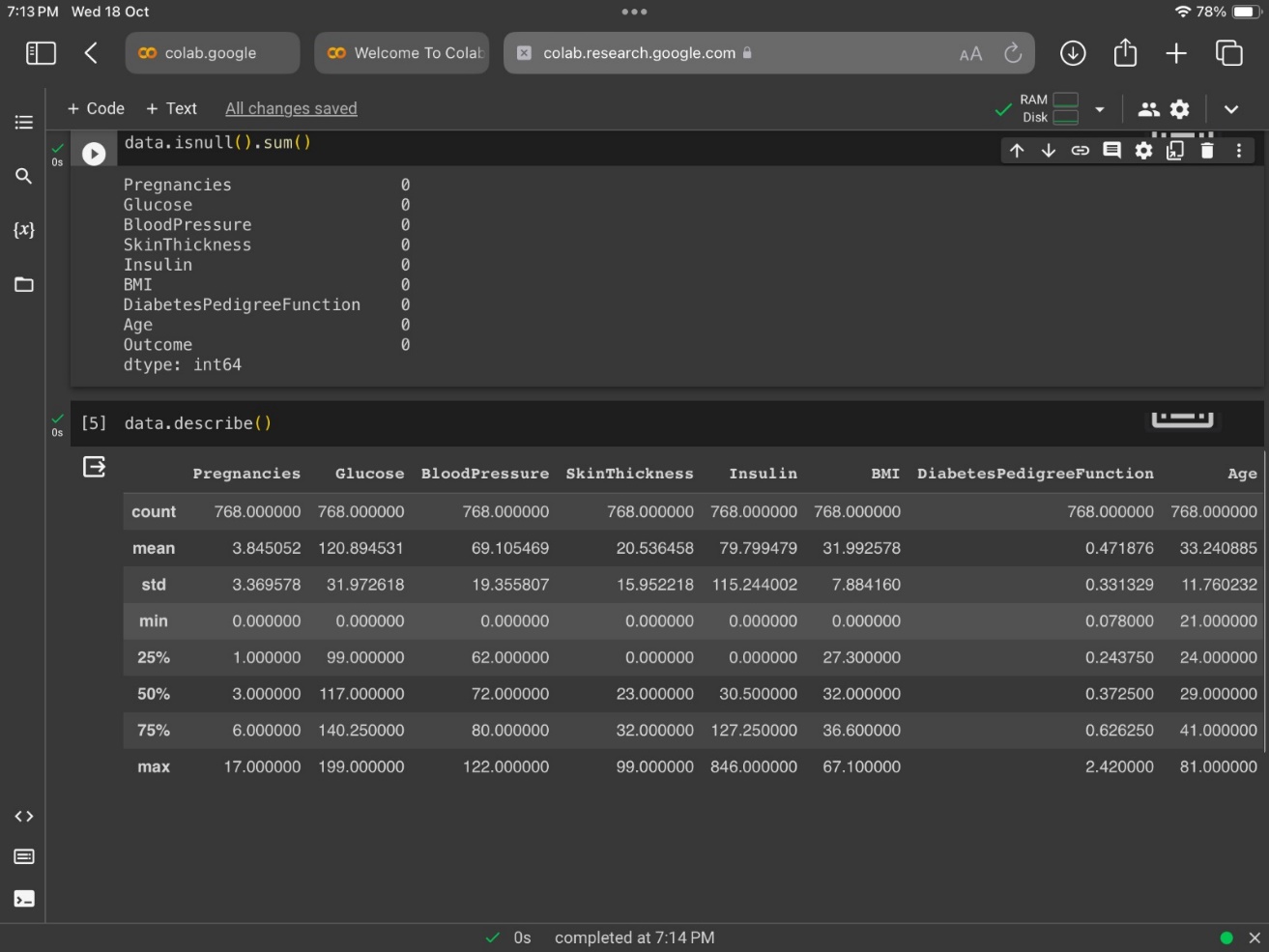
This topic describes how to import data to an existing dataset in a LabKey study. Authorized users can add data a row at a time, or import data from spreadsheets or other data files. Imported data must match the structure of the dataset into which it is imported.

The following import methods are available:

* [Insert Single Row](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#single) - Appends a single row of data to the dataset.
* [Bulk Import from File](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#file) - Append or [merge](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#merge) multiple rows from a file.
* [Bulk Import via Copy-and-Paste](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#paste) - Append or [merge](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#merge) multiple rows from a Copy-and-Paste.
* [Update and Merge](https://www.labkey.org/Documentation/wiki-page.view?name=uploadStudyData#merge) - For any bulk import method, existing rows may be updated.

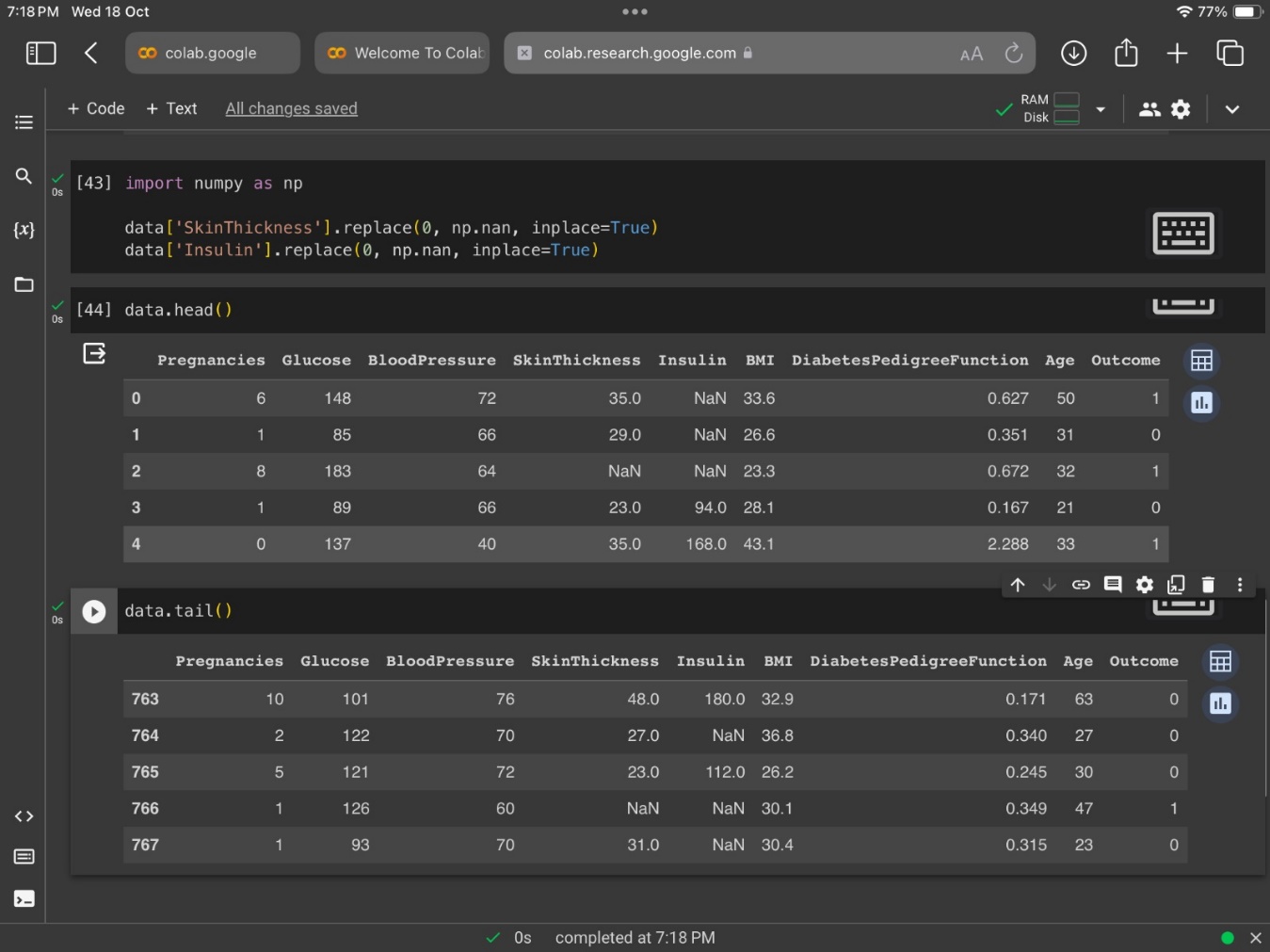






1. **Replacing 0 with NaN:**

we first import the necessary libraries, Pandas and NumPy. Then, we create a sample dataframe with three columns, 'A', 'B', and 'C', and replace the 0 values with NaN using the replace() method.



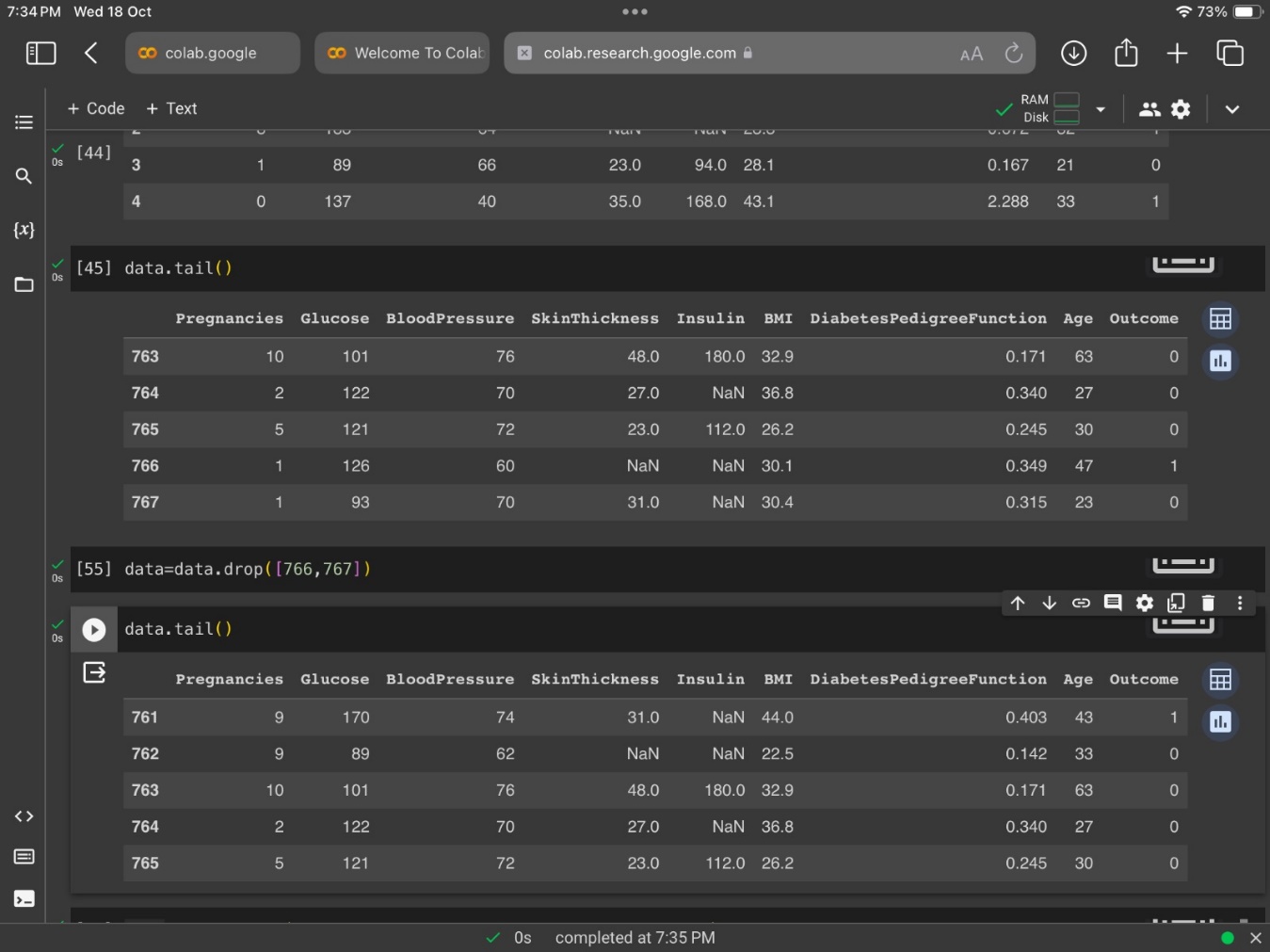
1. **Dropping last 2 rows to improve interpolation**

**Method 1:** Using [**Dataframe.drop()**](https://www.geeksforgeeks.org/python-delete-rows-columns-from-dataframe-using-pandas-drop/).

We can remove the last n rows using the drop() method. drop() method gets an inplace argument which takes a boolean value. If inplace attribute is set to True then the dataframe gets updated with the new value of dataframe (dataframe with last n rows removed).

**Method 2:**Using[**Dataframe.iloc[ ]**](https://www.geeksforgeeks.org/python-extracting-rows-using-pandas-iloc/)**.**

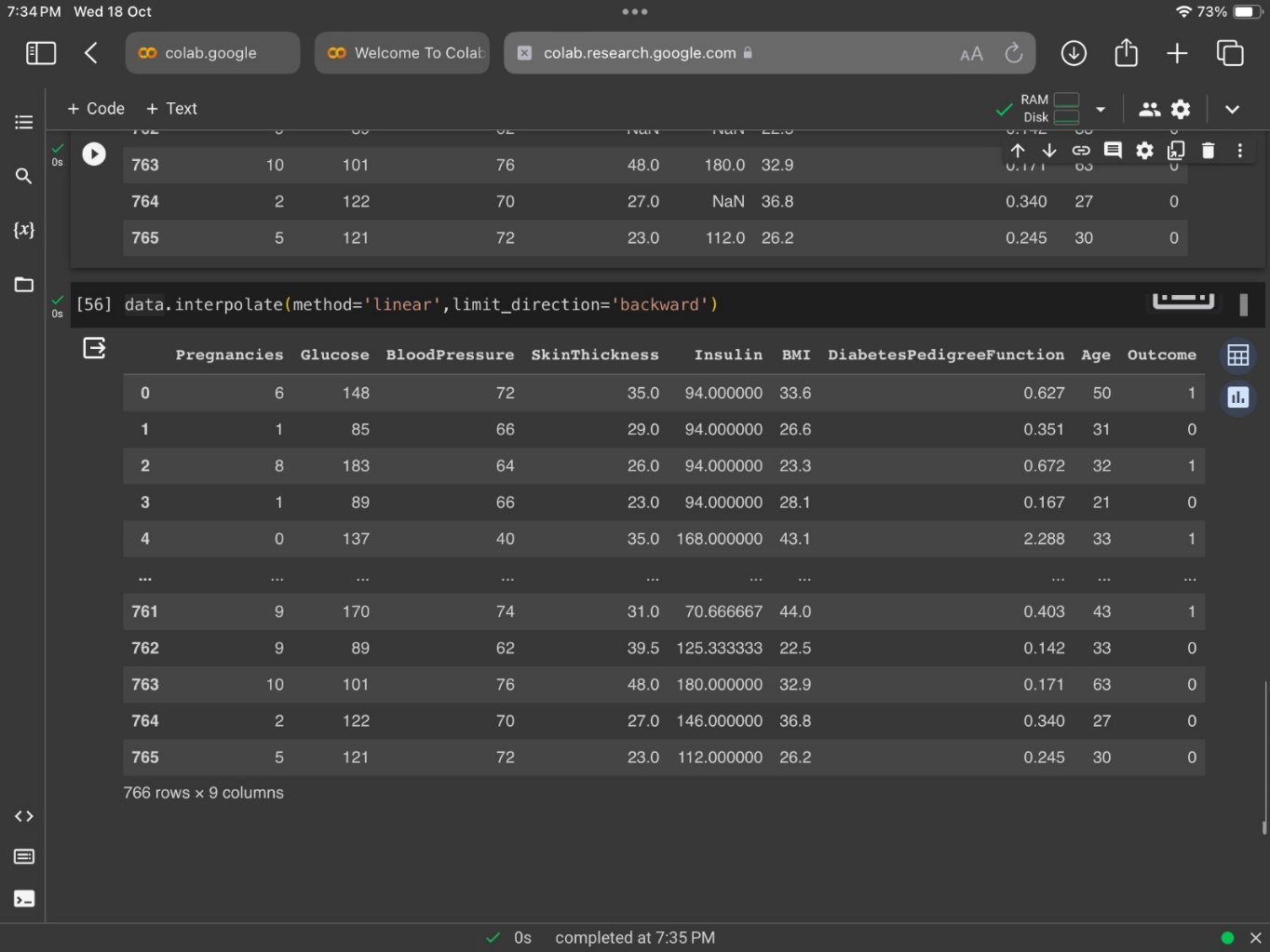
Thismethod is used when the index label of a data frame is something other than numeric series of 0, 1, 2, 3….n or in case the user doesn’t know the index label.



1. **Interpolating the missing values**

Linear Interpolation simply means to estimate a missing value by connecting dots in a straight line in increasing order. In short, It estimates the unknown value in the same increasing order from previous values. The default method used by Interpolation is Linear.

This beginner’s tutorial is about interpolation. Interpolation in Python is a technique used to estimate unknown data points between two known data points. In Python, Interpolation is a technique mostly used to impute missing values in the data frame or series while preprocessing data. You can use this method to estimate missing data points in your data using Python in Power BI or machine learning algorithms. Interpolation is also used in Image Processing when expanding an image, where you can estimate the pixel value with the help of neighboring pixels.



1. **Analysing the dataset**

Data analysis refers to the process of manipulating raw data to uncover useful insights and draw conclusions. During this process, a [data analyst or data scientist](https://online.hbs.edu/blog/post/data-analytics-vs-data-science) will organize, transform, and model a dataset.

Organizations use data to solve business problems, make informed decisions, and effectively plan for the future. Data analysis ensures that this data is optimized and ready to use.

Some specific types of data analysis include:

* Descriptive analysis
* Diagnostic analysis
* Predictive analysis
* Prescriptive analysis

Regardless of your reason for analyzing data, there are six simple steps that you can follow to make the data analysis process more efficient.

