1. Get familiar with the Koalas documentation: <https://koalas.readthedocs.io/en/latest/>
2. Walk through the live example: <https://mybinder.org/v2/gh/databricks/koalas/master?filepath=docs%2Fsource%2Fgetting_started%2F10min.ipynb>
   1. Run all cells in the notebook
3. Provide report with answers to the following questions:
   1. How to replace NULL in a Koalas Dataframe with specific value?  
      [**fillna**](https://koalas.readthedocs.io/en/latest/reference/api/databricks.koalas.DataFrame.fillna.html#databricks.koalas.DataFrame.fillna)([value, method, axis, inplace, limit])
   2. How to quickly collect stats for Pandas/Koalas Dataframe?  
      [**describe**](https://koalas.readthedocs.io/en/latest/reference/api/databricks.koalas.DataFrame.describe.html#databricks.koalas.DataFrame.describe)([percentiles])  
      import pandas\_profiling  
      profiled\_report = pandas\_profiling.ProfileReport(df)

profiled\_report.to\_file('./profile\_report.html')

* 1. How to perform left outer join between Koalas dataframes df1 and df2 by id column?

left\_outer\_join = ps.merge(df1, df2, how='left', on=['id'])

* 1. What are use cases for Pandas Dataframe and for Koalas Dataframe?  
     Pandas DF is suited for smaller scale computations on a single computer.  
     Koalas DF provides Spark distributed compute functionality over Pandas DF, and is suited for large scale data manipulations.
  2. What Koalas Dataframe’s Index is used for? What is Spark analogue of Koalas Dataframe’s Index (if any)?  
     1. Identifies data (i.e. provides metadata) using known indicators, important for analysis, visualization, and interactive console display.
     2. Enables automatic and explicit data alignment.
     3. Allows intuitive getting and setting of subsets of the data set.

In Spark we might use zipWithIndex() or zipWithUniqueIdfor the same purposes