Practice Quiz (Ungraded) - Statistics and API usage on Spark

Pregunta 1:

1. Why is it useful to compute statistical moments?

   ● Statistical moments tell us a lot about the value distribution of different features (table columns).

   ○ Derived values of statistical moments are used as input for downstream machine leaerning algorithms

   ✓ **Correcto**
   Correct

Pregunta 2:

2. Which API should I use? RDD, DataFrame or SQL?

   ● You should always use the DataFrame or SQL API because it makes use of the Tungsten and Catalyst optimizers resulting in faster running code and less development effort. RDDs exists under the hood and are only useful if a problem can't be expressed using the DataFrame or SQL API

   ○ RDD is always the preferred way to go because applying functional programming is the way to make any program run fastest. The DataFrame or SQL API are only for less skilled users until they make their way through to use RDDs

   ✓ **Correcto**
   Correct