Spark Code Skeleton

Spark + RDDs

```
from pyspark import SparkConf, SparkContext
import sys
assert sys.version_info >= (3, 5) # make sure we have Python 3.5+

# add more functions as necessary

def main(inputs, output):
    # main logic starts here

if __name__ == '__main__':
    conf = SparkConf().setAppName('example code')
    sc = SparkContext(conf=conf)
    sc.setLogLevel('WARN')
    assert sc.version >= '3.0'  # make sure we have Spark 3.0+
    inputs = sys.argv[1]
    output = sys.argv[2]
    main(inputs, output)
```

Spark + DataFrames

```
import sys
assert sys.version_info >= (3, 5) # make sure we have Python 3.5+

from pyspark.sql import SparkSession, functions, types

# add more functions as necessary

def main(inputs, output):
    # main logic starts here

if __name__ == '__main__':
    inputs = sys.argv[1]
    output = sys.argv[2]
    spark = SparkSession.builder.appName('example code').getOrCreate()
    assert spark.version >= '3.0' # make sure we have Spark 3.0+
    spark.sparkContext.setLogLevel('WARN')
    sc = spark.sparkContext
    main(inputs, output)
```

Updated Fri Aug. 25 2023, 15:55 by ggbaker.