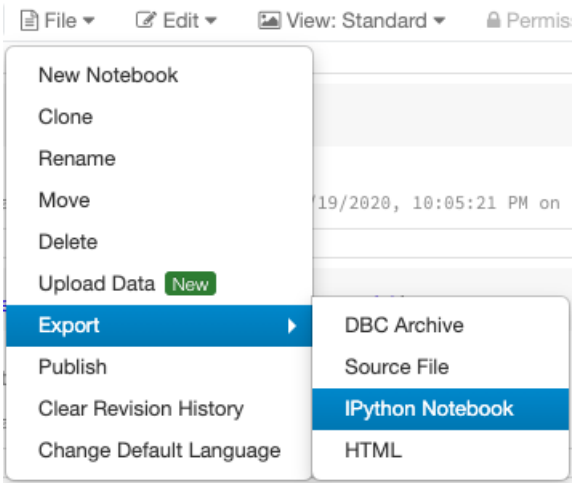
DS 610 Week 7 Assignment
Big Data Analytics

Due Date:

Please Note: As you will be working on Databricks console for this assignment, please submit the IPython Notebook [File-> Export-> IPython Notebook]. Use Markdown. Submissions in the form of screenshots / word documents or in any other format will **NOT** be evaluated. Please prefix your file name with your first name like below
name_ds610_assignment_07.ipynb



1. Using SparkSession and the file *ContainsNull.csv,* explain the significance of *how* and *thresh* arguments in *drop()* function.

2. Using SparkSession and the file *ContainsNull.csv,* fill the null sales values with the minimum sales value.

Output:

```
+----+-----+-----+
|  Id| Name|Sales|
+----+-----+-----+
|emp1| John|345.0|
|emp2| null|345.0|
|emp3| null|345.0|
|emp4|Cindy|456.0|
+----+-----+-----+
```

3. Using SparkSession and the file *appl_stock.csv,* show the unique trade years in descending order with the output column name as shown below.

Output:

```
+----+
|year|
+----+
|2016|
|2015|
|2014|
|2013|
|2012|
|2011|
|2010|
+----+
```

4. Using SparkSession and the file *appl_stock.csv,* show the average trade volume for each year with the output column names and values as shown below.

Output:

```
+----+----------------+
|year|Final Avg Volume|
+----+----------------+
|2010|  149,826,316.67|
|2011|  123,074,741.67|
|2012|  131,964,204.40|
|2013|  101,608,700.00|
|2014|   63,152,730.56|
|2015|   51,837,886.90|
|2016|   38,415,362.30|
+----+----------------+
```

5. What limitation of Hadoop (HDFS) does HBase overcome?

A) backup plans
B) transaction support
C) reliability
D) scalability

6. What does the C in the CAP Theorem of Big Data stand for?

A) columnar
B) capability
C) collated
D) consistency

7. Which type of Big Data Architecture does HBase use?
A) document store
B) column family
C) key / value store
D) graph database

8. What HBase construct is most similar to a database in a RDBMS?
A) version
B) namespace
C) row key
D) table collection

9. Besides 'default' what namespace does HBase come with preloaded?
A) info
B) meta
C) hbase
D) info_schema

10. What command adds a new row to an HBase table using the shell?
A) move
B) put
C) insert
D) create

Thank you.