It seems we forgot to attach our input files and MapReduce screenshots in the hw submission. Resending all files to add clarity.

Readme file:

Steps

1. Add the btc.txt, eth.txt and ltc.txt files to Dumbo and HDFS.

2. Run sentiment analysis using NEWS API script to provide a sentiment score to each news article using python script test.py and save new file as data.csv [Data Profiling]

3. Run mapper.py and reducer.py with input file named data.csv to get output file part-00000 (this file is later renamed to news.txt) and add it to Dumbo and HDFS [Data Profiling and MapReduce job]

4. Run hive commands in hive\_commands.txt to create tables using join (our main join table is Date-sentiment score-btc price-eth price-ltc price) and some other aggregate queries as described in the hive commands file. (Our database is jda448\_project. It has all the original tables and join tables for reference) [Analytics]