Exam for Machine Learning Python Lab

Consider the file provided with the assignment and perform the tasks described below.
The solution must be produced as a Python Notebook, assuming that the dataset is in the same folder as the notebook. The notebook must include appropriate comments and must operate as follows:

1. Read the dataset, show its head, shape and description

2. Eliminate totally null columns and totally null rows, eliminate columns with less than 1/3 of non null values; fill the remaining `NaN` values with the mean of the column **2pt**

3. Drop `Time`, convert `Date` from *string* to *datetime* and group by Date using mean as aggregate function **2pt**

4. Preparation of the boolean matrix **3pt**
   - Discretise *continuous values* with *two bins*, *kmeans* strategy and *onehot-dense* encoding
   - Discretization/encoding generates 0/1 values; convert the binary values obtained into boolean, as requested by *Apriori*

5. Set the names of two columns generated by the discretisation of each attribute A to A_low, A_high (with discretisation/one-hot-encoding, each original column generates two columns, the first is for the *low* values, the second for the *high* values) **3pt**

6. find a value of `min_support` such that the Apriori algorithm generates at least *8* frequent itemsets with at least *2* items, output the result **3pt**

7. find the minimum metric threshold such that at least 100 association rules are extracted from the frequent itemsets found and show the metrics used and the best 10 rules by descending confidence and support **3pt**

Quality of the code **4pt**
- Include appropriate comments with reference to the numbered requirements
- Useless cells, pieces of code and non-required output will be penalised
- Remove the code you use for testing and inspecting the variables during the development
- Naming style of variables must be uniform and in English
- Bad indentation and messy code will be penalised
- Non generalised solution, such as three sequential statements with the same kind of operation instead of a loop, will be penalised

Additional directions: the assignments not compliant with the rules below will not be considered:
- The *notebook name* must be in lowercase according to the pattern `workplacecode_youremailusername.ipynb` with underscore instead of dot inside the email username. E.G. if your email is `mario.rossi45@studio.unibo.it` and you are sitting in the workplace `lab4_023`, the notebook filename will be `lab4_023_mario_rossi45.ipynb`
- The solution must directly access the data in the same folder of the notebook, the name of the file must be the same as the file provided. If the notebook is developed using *Google Colab*, the code must be able to work also out of the Google Colab environment without any change.
- Check carefully that your file is correctly stored as a python notebook. Upload the notebook only to `http://eol.unibo.it` in the activity specified by the teacher, any other way of submitting the notebook will be ignored.
- Cooperative work will be heavily sanctioned both for the giver and the receiver of the copy.
- The candidate can freely access any kind of materials.