#1 - *Column #1 : SID* - The columns "SID", "Number" and "Name" have the same function, distinguish a cyclone from another. I suggest that we keep only the column "Number".

2 - *Column #6 : NAME* - In my opinion this column should be retreated. Indeed, it has too much information in one column : the exact day and time.

I suggest that from this column we create 2 different ones :

- Date : Should we keep only the month ?

- Time : Should we transform the exact time into "Morning", "Afternoon", "Evening" and "Night" ?

3 - *From Column #9 to 10 : LAT and LON* - As we already have the location with the column "Subassin", should we keep also the "LAT" and "LON", this information seems to be too precise for me...

4 - *Column #13 : WMO_AGENCY* - Does the reporting agency have an impact on a cyclone severity ? Probably not, that's why I suggest to delete it.

5 - *Column #17 : IFLAG* - Does the source of each agency's report have an influence on a cyclone severity ? Probably not...

6 - *Column #18 : USA_AGENCY* - This column deals once again with the source of the data. I don't think it has an impact on a cyclone severity.

7 - *Column #19 : USA_ACTIF_ID* - We already have this info : Bassin, number of the storm and the year in previous columns.

8 - *From Column #20 to #21: USA_LAT and USA_LON* - Using the latitude and longitude is too precise. Moreover, we won't use the local measures (USA, Tokyo, etc.) , only those contained in the columns 9 and 10 (LAT and LON).

9 - *Column #22 : USA_RECORD* - This column just give more details about the wind and the pressure column. I don't think we should keep it. Moreover, only the USA agency has this column, not the other agencies. Wind speed and pressure are enough.

10 - *Column #23 : USA_STATUS* - The status of the cyclone is already mentioned in the column number 8 (Nature). Although this classification seems to be more precise, the USA_STATUS is not mentioned for every cyclone. So we should delete it.

11 – *From column #24 to #25 : USA_WIND and USA_PRES* - We will keep the information about the Wind speed and pressure form the columns 11 and 12 (WMO_WIND and WMO_PRES). Not from the local agencies column.

12 - *Column #26: USA_SSHS* - This column is interesting but probably too close to the column number 145, "TD9636_STAGE", the one that the ML project must predict. I think we should delete it. But I'm not sure.

13 - *From Column #27 to #38* - This information about the wind radii maximum extent in each quadrant is not given by every agencies, so we should delete it. Moreover it seems to be a more detailed version of the wind speed...

14 - *From Column #39 to #40: USA_POCI and USA_ROCI* - Those information are said to be "NOT BEST TRACKED", and concern only the USA agency. We should delete it.

15 - *From Column #41 to #144 and from #146 to #172* - The following columns concern the measures of the local agencies. They will be deleted.