Approach to solve this dataset

What data-pre-processing / data cleaning ideas really worked? How did you discover them?

1. In this competition, we have given two different datasets. One is Visitor Log Dataset which contains all the information like Product id, user id , activity of user , web client id when someone visit a particular website . In another dataset user table , we have User Id, Sign Up date of user and User segment like A,B,C
2. Total No. of records in Visitor Log Dataset is 65,88,000 with 9 columns and Total No. of records in User Table dataset is 34,050 with 3 columns
3. First I removed those records where user ID was not given.
4. In Date and time column, three types of formats were there. First dates were given in date time format . second dates were given in unix format and third there were null values in the datetime column . I created two different dataset by filtering records with null values and not null values then from not null values I created two different dataframe where datetime was given in datetime format and datetime was given in unix format. After all the values given in datetime column converting them in datetime format . I concatenate all these dataframe into one dataframe.
5. For Filling NA values in DateandTime Column . I sorted the data as per [UserID,WebclientID,ProductID,VisitDateTime] then then filled NA values as previous given datetime .
6. For Filling NA values in ProductID , I sorted data as per UserID -> Visit_Date_Time and used FFILL method to fill ProductID.

Which tools did you use to solve the problem?

1. I Used pandas to solve the problem.
2. The Challenging part was to iterate each and every user through whole dataset to get the desired filtered output. First I used pandas.apply method but it was very time consuming process Then I used groupby with dictionary and then map the data by userid which was the best way to solve these kind of problems .