The given data corresponds to the Internet surfing information for a famous web browser. Analyze the following problems with help of Python data science libraries.

Note: Ensure that your visualizations are easily readable.

1. Modify the dataset such that it only contains the following columns
   Hashed_ui, Country, Domain, Device, Screen_width, Screen_Height, DomainCategory, Installation_Day, and Pageviews
2. Remove the rows with missing or **b'unknown'** data in any of the columns.
   a. What is the number of rows in the resultant data?
   b. What is the percentage decrease in the number of unique users?
3. Impute the missing values. Find an appropriate replacement (if any).
4. Categorically encode the feature "DomainCategory".
5. What fraction of domains in the dataset belongs to the unclassified domain category? If this fraction is significant, remove such rows and again report
   a. the number of rows in the resultant data.
   b. the percentage decrease in the number of unique users?
6. Draw a histogram of [Top k; k is any positive integer]
   a. Most visited domains in terms of unique users
   b. Most visited domains in terms of the number of pages
   c. Most visited domains in terms of the average number of pages a user watches
7. Draw a histogram of
   a. Most visited domain categories in terms of unique users
   b. Most visited domain categories in terms of the number of pages
   c. Most visited domain categories in terms of the average number of pages watched by a user
   d. The number of users per country
   e. Country vs screen size
8. Are particular kinds of devices/ device sizes (screen dimensions) inclined towards particular domains/domain categories?
9. [Comparison of the popularity of Handheld Devices across countries] Any particular kinds of devices/device sizes (screen dimensions)?
   a. Granular analysis
      i. By Screensize (Length/Width)
      ii. By Carrier
      iii. [Market Capture] By Company (Apple/Samsung/LG others..)
10. Create a pie chart for domain categories based on the number of domains they consist of. Choose to include a limited number of categories so that the chart is readable.
11. Create a pie chart for domain categories based on the number of unique users they serve. Choose to include a limited number of categories so that the chart is readable.
12. How diverse are the users in terms of the domain categories they watch? Think of a relevant metric to capture the diversity.
13. Are users from different countries inclined towards particular domains/domain categories?

14. Based on the screen's width and height, we can predict whether the user was using a mobile, laptop, or iPad. Are users inclined towards using this web browser through a particular device?
15. Can you apply a clustering algorithm on the users based on various given features? Report the performance of your algorithm using a relevant metric.
16. From the dataset, you can observe that people keep switching from one domain to another. Find the average number of domains a person explores in one surf.
17. Figure out the 10 most frequent pairs of domains explored consecutively.
18. Figure out the 10 most frequent pairs of domain categories explored consecutively.
19. Correlation with the usage (PageViews) v/s (ScreenSize)

#You can add experiments of your own to extract more insights from this dataset.