Solve Using Numpy, Pandas, Matplotlib and seaborn

Case study 1:
Dataset Description: The file consists of start-ups investment details.

1. Read the given comma separated file as dataframe.
2. List out all columns names.
3. Create a dataframe with numerical columns.
4. Create a dataframe with categorical columns.
5. Get summary on the data and draw inferences if any.
6. Display duplicate rows.
7. For each column find out percentage of missing values.
8. Find count of ‘name’ in each ‘country_code’.
9. What is the percentage of the companies which have status ‘acquired’ and ‘operating’?
10. Create a column ‘category_list_count’ having count of category lists.
11. Find total ‘fundings’ for each country_code.
12. Find average ‘fundings’ for each country_code.
13. Find average ‘fundings’ in each region.
14. How many companies have got just 1 round of funding?
15. How many companies have ‘debt_financing’ above zero?
16. Create a column ‘homepage’ to store company name from ‘homepage_url’: For example: If url is http://www.waywire.com, name is waywire.
17. Find count of companies in each of the market.
18. Rename ' funding_total_usd ' to 'funding_total_usd'
19. For each row in column ‘funding_total_usd’, calculate actual – average value for each group ‘city’
20. What is average ‘funding_Total_used’ for each city?
21. Plot histogram/distribution of ‘funding_total_usd’ and provide insights if any.
22. What is maximum ‘funding_total_usd’ for each market status?
23. How many years it have been since each company was founded?
24. Visualize ‘grant’ distribution.
25. Visualize ‘debt_financing’ distribution.

Case Study 2:
Load practice.csv file as a data-frame and perform following operations on the data-frame

1. Display all columns
2. create numerical and categorical columns list
3. display size of the data-frame
4. rename column MSSubClass -> SubClass, MSZoning -> Zones
5. display distinct values for Zoning, LotShape, LotConfig
6. display count of distinct values for Zoning, LotShape, LotConfig
7. max, min of column YearBuilt

8. create a new column “year_diff’. This will be holding difference of current year and YearBuilt
9. display distinct MSZoning for each OverallQual
10. what is maximum LotArea where BsmtExposure = Mn?
11. Sort dataframe based on following columns and orders: MSSubClass; ascending, YearBuilt; descending
12. What is average OverallQual.
13. Group by YearBuilt and find maximum OverallQal
14. Load the data_1.csv again with MSSubClass as new index
15. Convert LotArea as numpy array
16. In column MasVnrArea replace 0 with -1
17. Check if there is/are any Null values (NaN) in the data given
18. Display percentage of missing values in each column if any
19. Select records where LotConfig is Inside
20. Make a new dataframe with only numeric columns
21. Make a new dataframe with only factorial/string columns
22. Drop column ExterQual
23. Group data on LotShape and find average LotArea