Case Studies Using Numpy, Pandas, matplotlib & seaborn

*Case Study 1:*

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

1. Read the given comma separated values as dataframe (investments.csv)
2. List out all column names.
3. Create a dataframe with numerical columns.
4. Create a dataframe with categorical columns.
5. Get a summary on the data and draw inferences if any.
6. Display duplicate rows.
7. For each column find out the 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’ ‘operating’?
10. What is the percentage of the companies which have status ‘acquired’ acquired?
11. Filter records having missing values in column ‘year\_founded’.
12. Create a column ‘category\_list\_count’ having count of category lists.
13. Find average funding\_total\_usd for each country\_code.
14. Find total funding\_total\_usd for each country\_code.
15. Find average funding\_total\_usd in each country\_code and region.
16. How many companies have got just 1 round of funding?
17. Perform mapping on status column; acquired -> A, operating -> O and closed -> C.
18. How many companies have ‘debt\_financing’ above zero?
19. Create a column ‘homepage’ to store company name from ‘homepage\_url’: For example: If url is <http://www.waywire.com>, name is waywire.
20. Find the count of companies in each of the markets.
21. Find the count of companies in each of the markets and store the new column ‘cnt\_name’ in the original dataframe.
22. Rename ' funding\_total\_usd ' to 'funding\_total\_usd'
23. Fill missing values in column ‘city’ with ‘other\_city’
24. For each row in column ‘funding\_total\_usd’, calculate actual – average value for each group ‘city’
25. Normalize ‘‘funding\_total\_usd’ at country level.
26. What is the average ‘funding\_total\_usd’ for each city?
27. Plot histogram/distribution of ‘funding\_total\_usd’ and provide insights if any.
28. What is maximum ‘funding\_total\_usd’ for each market status?
29. How many years has it been since each company was founded?
30. Visualize ‘grant’ distribution.
31. Visualize ‘debt\_financing’ distribution.
32. Display proportion of companies status.
33. How many US states are available?
34. create column ‘cmt\_address’ by joining country code, state code, region and city.
35. select columns with underscore in their names.