Module-3

1. How does cleaning data facilitate easier storage, search, and reuse.Describe the process of identifying values for data cleanup.
2. Describe how to replace the short headers with longer. Consider the following mn.csv and mn-headers.csv files.

sl.no	HH1	HH2	LN	MWM10H	MWM10M	MWM11H	MWM11M	MWB1M	MWB1Y	MWB2	MWB3	MWB4	MWB5	MWB7	TN12_1	TN12_2	TN12_3	TN12_4
`	1	17	1	17	59	18	7	5	1984	29	Yes	Higher	31	NA	NA	NA	NA	NA
2	1	20	1	17	32	17	42	5	1976	37	Yes	Higher	31	NA	NA	NA	NA	NA
3	2	1	1	10	37	10	52	2	1973	41	Yes	Primary	17	Able to re	NA	NA	NA	NA

Name	Label
HH1	Cluster number
HH2	Household number
LN	Line number
MWM10H	Start of interview - Hour
MWM10M	Start of interview - Minutes
MWM11H	End of interview - Hour
MWM11M	End of interview - Minutes
MWB2	Age of man
MWB3	Ever attended school
MWB4	Highest level of school attende
MWB5	Highest grade completed at tha
MWB7	Can read part of the sentence
MLS2	Estimation of overall happiness
MLS3	Satisfaction with family life
MLS4	Satisfaction with friendships
MLS5	School attendance during the c
MLS6	Satisfaction with school
MLS7	Satisfaction with current job
MLS8	Satisfaction with health
MLS9	Satisfaction with current reside
MLS10	Satisfaction with treatment by
MLS11	Satisfaction with appearance
MLS12	Satisfaction with life overall
MLS13	Satisfaction with current incom
MLS14	Life satisfaction in comparison
MLS15	Life satisfaction expectation or
TN12_1	Person 1 who slept under net
TN12_2	Person 2 who slept under net
TN12_3	Person 3 who slept under net
TN12_4	Person 4 who slept under net
TN2	Number of mosquito nets
PSU	Primary sampling unit
strata	Stratum

3. Explain what an f-string is in Python and provide an example of how it can be used to format a string with variables.

4. Write Python code and explain how to print the current date and time in the format YYYY-MM-DD HH:MM:SS

5. Given a list colors = ["red", "green", "blue"], write code to output these colors as a comma-separated string.

6. Write Python code to format the following dictionary data in order to get the desired output:

```
data = {
    'price': 199.99,
    'quantity': 1000,
    'discount_rate': 0.15
}
```

output:

Price: $199.9900

Quantity: 1,000

Discount Rate: 15.000%

7. Illustrate the purpose of np.where(np.abs(z_scores)>threshold)[0]. What does it identify in the dataset?

8. Compare and contrast using Python's built-in set and NumPy's unique method for identifying unique elements in datasets.

9. Outline the purpose of using fuzz.ratio() and fuzz.partial_ratio() functions from the fuzzywuzzy library. What do these functions measure in terms of string similarity?

10. Illustrate the role of regular expressions. Describe the difference between re.findall() and re.search() methods in terms of their functionality and the type of results they return.

11. Discuss the significance of normalizing scores to a range between 0 and 1 and how it facilitates fair comparison of team performance across different categories.

12. Develop a Python script that interacts with an SQLite database to store and retrieve your data.

Module-4

1. Explain the steps to determine the number of sheets and their names and display the sheet_row_values in the workbook using the xlrd library, and describe how you would proceed with the agate library to explore and analyze the data further.
2. Make use of agate data analysis library to create agate table and explain how it will access the data according to its documentation with example.
3. Utilize different agate data analysis library functions to perform different operations on the agate table.
4. Elaborate on joins and Compare different methods to join datasets with examples for each.
5. Construct a python script to provide insights into the distribution and characteristics of the data within each category using aggregate functions.
6. Construct print bars using python script to visualize the aggregated data by taking sample example.
7. Identify and explain the different tools that can helps in digital storytelling and data presentation.
8. Develop a python script for visualizing the dataset in bar chart using matplotlib.pyplot library.
9. Compare and contrast the concepts Ghost and GitHub Pages and Jekyll.