CMP404

Introduction to Data Science

Assignment 5

TASK 1

- 1. Create 3 list viz. productId, Product Name, Product Price with 5 elements each.
- Add this element to create a DataFrame.
- 2. Extract one by one column and display separately from above dataframe.
- 3. Display separately each and every row from an above dataframe. (use loc or iloc function)
- 4. Display product name from 3 row using loc function.
- 5. Display price located in 2 row using iloc function.
- 6. Display 2nd to 4th row data excluding price.
- 7. Display 1 to 3rd row price data only.
- 8. Change marks in python column to 41 for students who have scored less than 40.
- 9. Add 2 marks to all the subjects marks. If any of the marks exceeds 100 change it to 100.
- 10. Display students who have scored more than 50 marks in python and more than 60 marks in data Science.
- 11. Display students records whose name starts with letter S.
- 12. Display students records whose name ends with letter a.
- 13. Calculate total and percentage for all the students.
- 14. Calculate Status (PASS / FAIL) for the Students.
- 15. Calculate percentage based grade value (per>=90 then A-grade, per>=80 then B-grade, per>=70 then C-grade, per>=60 then D-grade, per>=50 then E-grade, per<50 then F2grade)
- 16. Also calculate overall columns total, maximum, minimum and standard deviation for each numeric columns

Convert the above dataframe to excel, json and html.

- 17. Display meta-data for each column for the above dataframe.
- 18. Convert the above DataFrame to list within a dictionary.
- 19. Convert this dictionary again to a separate dataframe by adding 500 to each element in product price column.

Task 2

Using the <u>SF Salaries Dataset</u> from Kaggle! ,complete the tasks outlined in bold below

- 1. Import pandas as pd.
- 2. Read Salaries.csv as a dataframe called sal.
- 3. Check the head of the DataFrame.
- 4. Use the .info() method to find out how many entries there are.
- 5. What is the average BasePay?
- 6. What is the highest amount of OvertimePay in the dataset?
- 7. What is the job title of JOSEPH DRISCOLL? Note: Use all caps, otherwise you may get an answer that doesn't match up (there is also a lowercase Joseph Driscoll).
- 8. How much does JOSEPH DRISCOLL make (including benefits)?
- 9. What is the name of highest paid person (including benefits)?
- 10. What is the name of lowest paid person (including benefits)? Do you notice something strange about how much he or she is paid?
- 11. What was the average (mean) BasePay of all employees per year? (2011-2014)?
- 12. How many unique job titles are there?
- 13. What are the top 5 most common jobs?
- 14. How many Job Titles were represented by only one person in 2013? (e.g. Job Titles with only one occurence in 2013?)
- 15. How many people have the word Chief in their job title?
- 16. Is there a correlation between length of the Job Title string and Salary?