WEEK 2 TASK- COMPANY SALES DATASET

Dataset: https://pynative.com/wp-content/uploads/2019/01/company_sales_data.csv

1.Get familiar with colab
https://www.youtube.com/watch?v=6Xt6L1I5jSc&ab_channel=UnfoldDataScience

2. Make github account
https://www.youtube.com/watch?v=RGOj5yH7evk&ab_channel=freeCodeCamp.org

3. **Define** : Dictionaries, lists, tuples, set with an example
4. Read csv file
5. **Numpy:**

Q.1) Create an array of 10 fives.
Q.2) Create an array of all even integers between 10 to 50.
Q.4) Create a 4 x 4 matrix.
Q.5) Use numpy to generate an array of 25 random numbers sampled from a standard normal distribution.
Q.6) (a) matrix : ([[1 2 3 4 5] Output the given matrix using numpy.
[6 7 8 9 10]
[11 12 13 14 15]
[16 17 18 19 20]
[21 22 23 24 25]])

(b) by referring the above matrix output given matrix : ([[12 13 14 15]
17 18 19 20]
22 23 24 25]])
(c) Grab number '20' from the matrix.
(d) Output this array ([[2]
[7]
[12]])
(e) Get the standard derivation of the values in the matrix.
(f) Get the sum of all columns in the matrix.

6. **Pandas**:
1. Import pandas and check the version
2. Create a series from a list, numpy array and dict
input:

```
import numpy as np
mylist = list('abcedfghijklmnopqrstuvwxyz')
myarr = np.arange(26)
mydict = dict(zip(mylist, myarr))
```

3. Combine many series to form a dataframe

   input:
   import numpy as np
   ser1 = pd.Series(list('abcedfghijklmnopqrstuvwxyz'))
   ser2 = pd.Series(np.arange(26))
   get the items not common to both series A and series B
   ser1 = pd.Series([1, 2, 3, 4, 5])
   ser2 = pd.Series([4, 5, 6, 7, 8])

4. Get frequency counts of unique items of a series
   input:
   ser = pd.Series(np.take(list('abcdefgh'), np.random.randint(8, size=30)))

5. Stack two series vertically and horizontally
   input:
   ser1 = pd.Series(range(5))
   ser2 = pd.Series(list('abcde'))

6. Convert a series of date-strings to a timeseries
   input:
   ser = pd.Series(['01 Jan 2010', '02-02-2011', '20120303', '2013/04/04', '2014-05-05', '2015-06-06T12:20'])

7. Compute the euclidean distance between two series
   input:
   p = pd.Series([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])
   q = pd.Series([10, 9, 8, 7, 6, 5, 4, 3, 2, 1])
8. Import only specific columns from a csv file(first and second)
9. Use iloc(2 columns and 10 rows(any))
10. Find out the unique values in each column.

7. Check null values
8. Read and print the first and last five rows of the dataframe.
9.

Seaborn:

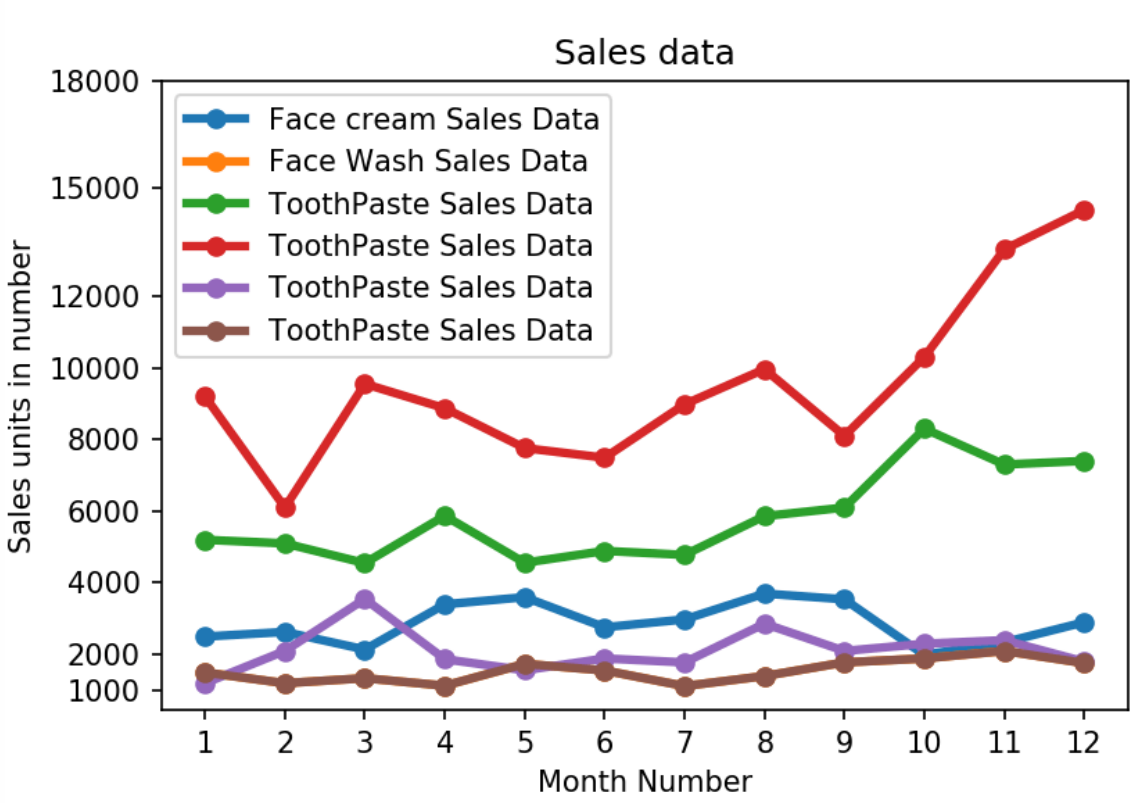
Q.1) Plot Jointplot and visualise the data.
Q.2) Plot Distplot , make kde to false and make bins to 30.
Q.3) Plot Scatterplot , try adding hue in it.
Q.4) Plot swarmplot and change color palette to ‘VIBGYOR’.
Q.5) Plot countplot and use hue parameter in it.
Q.6) Correlate the data and plot Heatmap on it and change its color scale to any you like.

Q.7) Plot a facetgrid and use hue in it. (Tricky one)
Q.8) Plot a pairplot.

Note: You can play with the parameters of the plots (for eg. bins,kde,color etc.)

10.
Matplotlib:

1. Show the relationship between Total profit and no. of month using Line Plot.
2. Read all product sales data and show it using a multi line plot. Display the number of units sold per month for each product using multi line plots. (i.e., Separate Plotline for each product).The graph should look like this.



3.
4. Plot a scatter plot between no. of month and toothpaste units sold.
5. Add a gridline to the scatter plot.
6. Plot a bar chart which displays the number of units sold per month for face cream and face wash. Add a separate bar for each product in the same chart.(with grid lines)
7. Plot the total profit of each month and show it using the histogram to see the most common profit ranges
8. Plot a pie chart for total sales data.

9. Read all product sales data and show it using the stack plot.

   Note:
   You have to plot a legend in each task.

11. Write about evaluation reports and types of it.

NOTE:
No task should change the default data that means you have to create new variables for each task.

Tools & Libraries You may need:
• Python
• Matplotlib, pandas, numpy, seaborn
• Colab

Due Date: 6th of August 2021, 8:00 PM
For any doubts contact the following mentors:

- Om Rajgure +91 7218126168
- Shailza Bansal +91 8955522295

After you're done with your work show it to us and then post it on LinkedIn by Tagging all 2 mentors and Cureya Team.