Assignment 6.2: Dataset Visualization

Gabi Rivera || 13Oct2022 || ADS501-01

In [2]:
```
import os
os.getcwd()
```

Out[2]:
```
'/Users/gabirivera/Desktop/MSADS2/ADS-501-01/Module 6/Code'
```

In [3]:
```
import pandas as pd
```

In [4]:
```
sp = pd.read_csv('ShampooSales.csv', sep = ',')
sp.head(10)
```

Out[4]:

	Month	1995	1996	1997
0	January	266.0	194.3	339.7
1	February	145.9	149.5	440.4
2	March	183.1	210.1	315.9
3	April	119.3	273.3	439.3
4	May	180.3	191.4	401.3
5	June	168.5	287.0	437.4
6	July	231.8	226.0	575.5
7	August	224.5	303.6	407.6
8	September	192.8	289.9	682.0
9	October	122.9	421.6	475.3

In [5]:
```
import klib
```

In [6]:
```
klib.missingval_plot(sp)
```

```
No missing values found in the dataset.
```

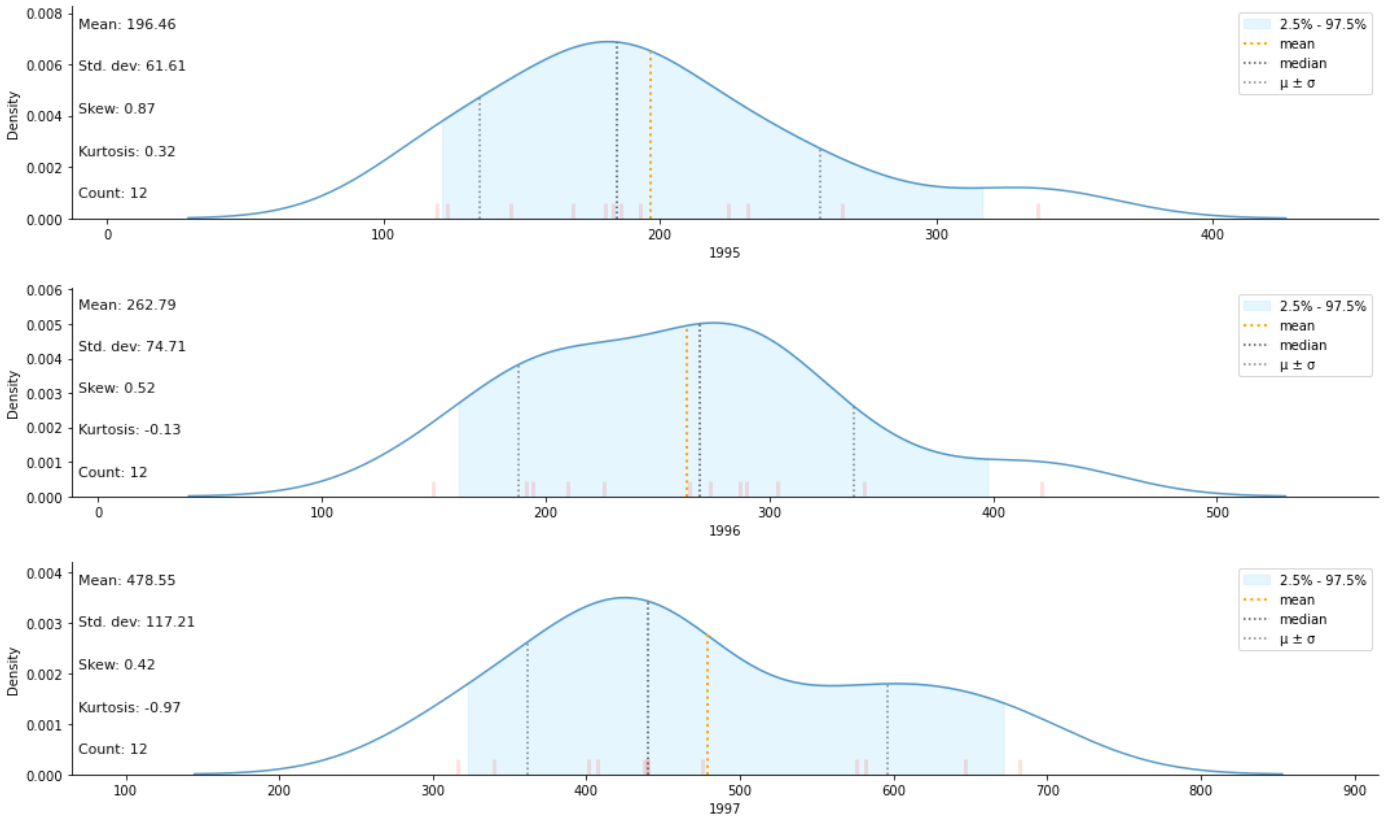
In [7]:
```
sp.describe()
```

Out[7]:

	1995	1996	1997
count	12.000000	12.000000	12.000000
mean	196.458333	262.791667	478.550000
std	61.606913	74.708007	117.212383
min	119.300000	149.500000	315.900000
25%	162.850000	206.150000	406.025000
50%	184.500000	268.900000	439.850000
75%	226.325000	293.325000	576.950000
max	336.500000	421.600000	682.000000

In [8]:
```
klib.dist_plot(sp)
```

Out[8]:
```
<AxesSubplot:xlabel='1997', ylabel='Density'>
```

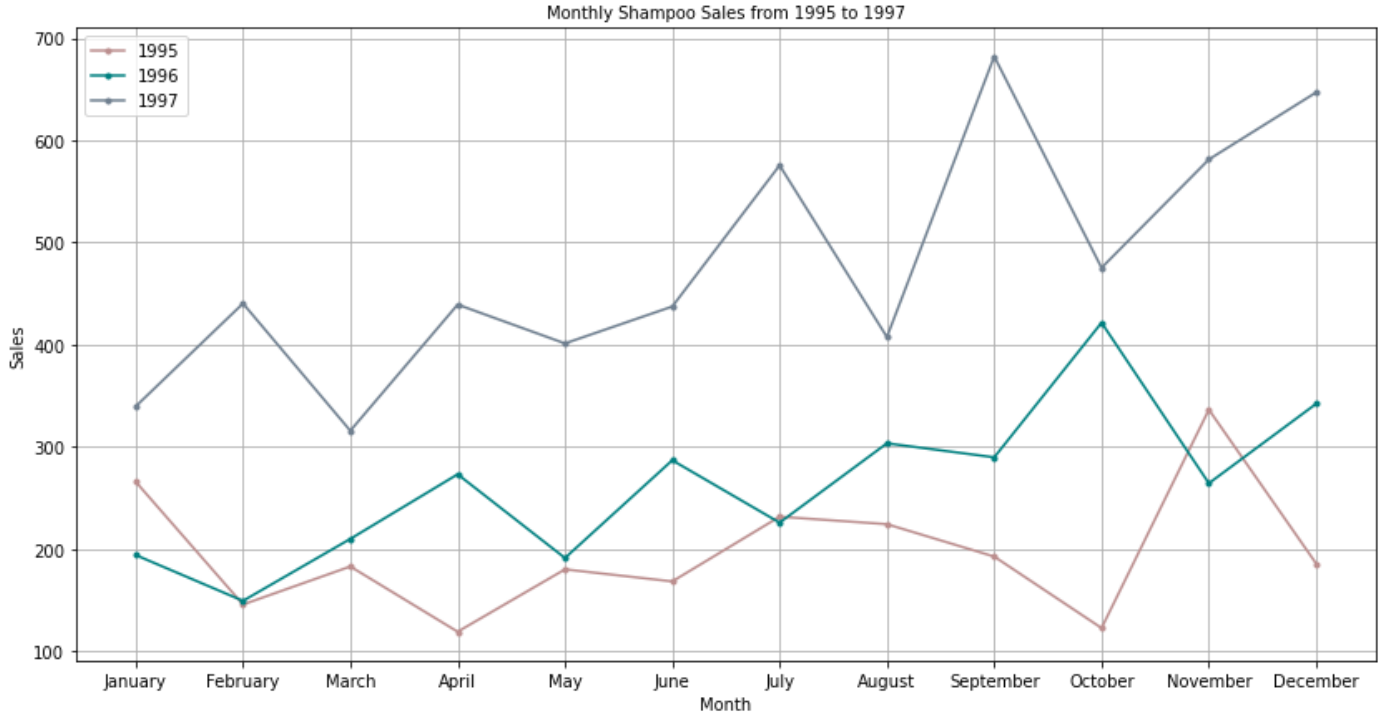


Direction: Create a line graph using the ShampooSales.csv file. Create your "x" axis using "Month", the "y" axis should be the sales, and then create a line for each year (1995, 1996, & 1997) to compare the sales per month by each year.

In [9]:
```
import matplotlib.pyplot as plt
```

In [15]:
```
plt.plot(sp['Month'], sp['1995'], color='rosybrown', marker='.')
plt.plot(sp['Month'], sp['1996'], color='teal', marker='.')
plt.plot(sp['Month'], sp['1997'], color='slategrey', marker='.')


plt.title('Monthly Shampoo Sales from 1995 to 1997', fontsize=10)
plt.xlabel('Month', fontsize=10)
plt.ylabel('Sales', fontsize=10)
plt.grid(True)
plt.legend(['1995', '1996', '1997'])
plt.rcParams['figure.figsize'] = [14, 7]
plt.show()
```



Interpretation:

Overall, 1997 shampoo sales are higher compared to 1995 and 1996 monthly shampoo sales. The relative increase in shampoo sales per year and the second half of the year might be driven by consumers' monthly prediction, shampoo cost response, increasing population, inflation, rising temperature, etc. On average, 1996 and 1997 monthly shampoo sales seem to have a cyclical pattern of decreasing after every other month. This might correspond to how long shampoo is consumed per household. 1995 almost had the same cyclical pattern but a slow dropped in sales is seen during the peak of the summer months (incline from June to October drop). This might be a response to the summer heat that year.