Introduction

Data visualization is a way to represent the datas/informations in graphical view. There has several type of insight to show this graphical view. These are: Chart, Column Chart, Bar Graph, Stacked Bar Graph, Stacked Column Chart, Area Chart. Dual Axis Chart, Line Graph, Mekko Chart, Pie Chart, Waterfall Chart, Bubble Chart, Scatter Plot Chart, Bullet Graph, Funnel Chart, Heat Map, etc.

Discussion

1st I read the daily visitors csv. And then stored in a new dataframe. Then found they have 40 unique Identifier and given a date column of 365(1 year). Then I compress the date to month. So total month is 12. And calculated MEAN. Then converted the rows(Date) to Columns and ID to rows.(Figure 1)

	2019-01-31	2019-02-28	2019-03-31	2019-04-30	2019-05-31	2019-06-30	2019-07-31	2019-08-31	2019-09-30	2019-10-31	2019-11-30	2019-12-31
UFY	62.064516	62.892857	62.967742	63.000000	63.806452	65.133333	65.580645	66.322581	66.000000	67.193548	66.733333	68.258065
XJT	78.419355	80.500000	80.000000	82.366667	81.806452	82.866667	83.935484	82.935484	83.733333	87.580645	84.833333	85.935484
XPE	78.258065	74.821429	78.225806	75.466667	76.129032	79.800000	73.290323	79.322581	75.500000	72.483871	71.166667	73.838710
AWF	140.193548	138.071429	135.870968	141.300000	138.258065	137.500000	143.354839	137.354839	140.066667	141.903226	138.600000	142.741935
AXM	58.451613	63.357143	60.774194	61.133333	63.096774	61.133333	62.709677	61.870968	64.300000	64.193548	64.200000	67.451613
BKI	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	32.419355	43.387097	35.766667	54.677419	74.000000	106.612903
XXO	93.483871	95.285714	94.677419	95.200000	96.612903	97.600000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
WRL	80.580645	76.642857	78.709677	75.100000	77.225806	76.800000	76.741935	78.516129	75.333333	75.935484	74.966667	74.354839
ZJB	0.000000	0.000000	0.000000	16.533333	29.838710	40.533333	32.290323	56.258065	42.266667	60.161290	52.566667	67.451613
PDT	252.580645	256.321429	260.903226	259.366667	276.387097	264.000000	267.967742	251.096774	279.133333	270.387097	283.266667	270.354839

(Figure 1: Top 10 rows of Daily visitors monthly)

The reason of making monthly is to read easily and make visualizations very easily. And it will be very easy to find the progress of each month.

Normal Seasonality visualisation of Mean by *Monthly* view

```python
import matplotlib.pyplot as plt
plt.figure()
plt.plot(df)
plt.show()
```
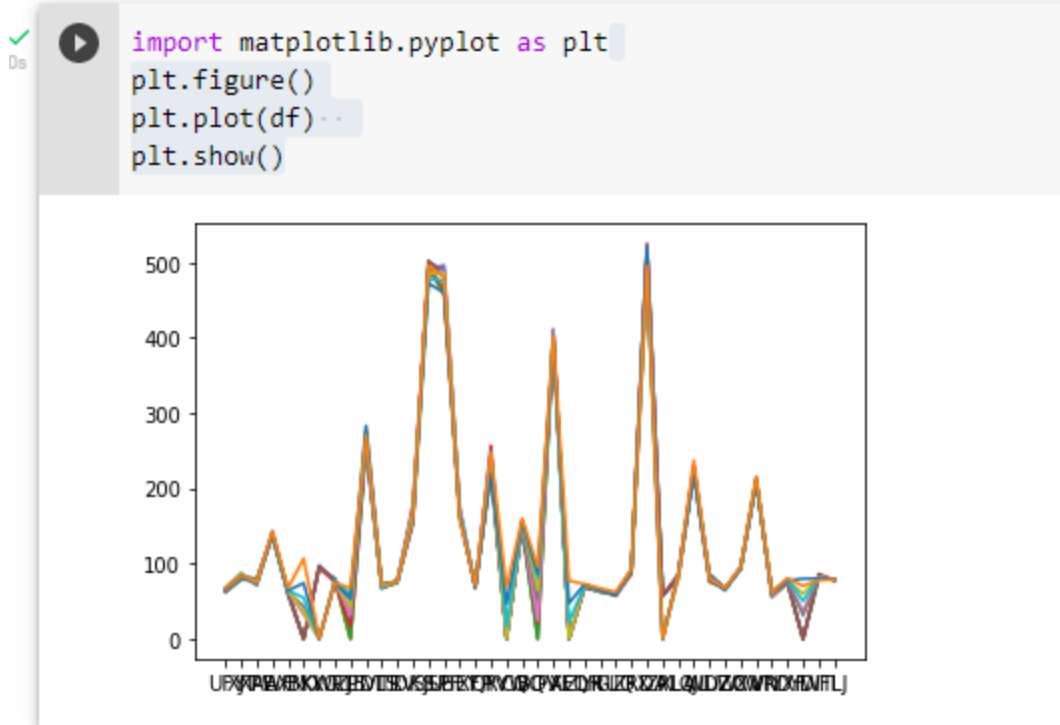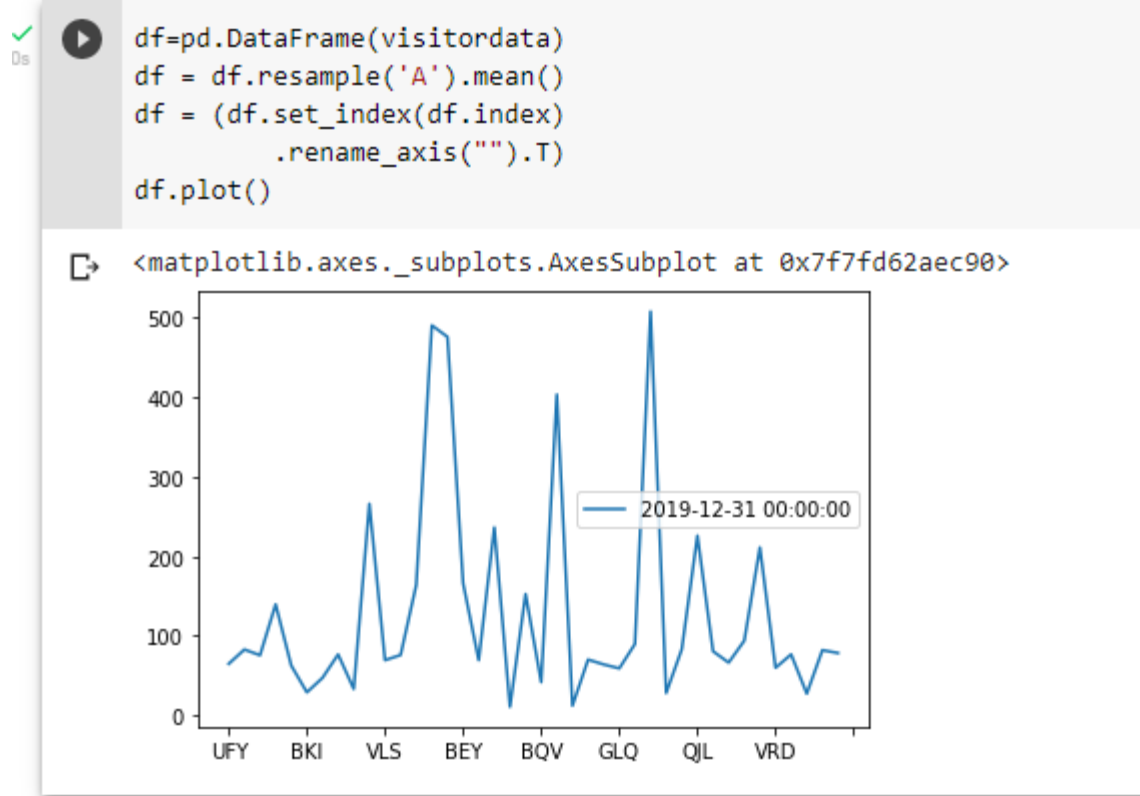

(Figure 2: Normal Seasonality visualization of Mean by Monthly view)

It will show all Venue visitors monthly average into this chart (Figure 2).

It is showing very ugly (Figure 2) but we can understand the progress of it. So, to fix it, I have used from 0 - 12 months (1 year) Average. So, now it will look beautiful and we can easily understand each venue Average (Figure 3).

*Averegare view of 12mos/1yr *

```python
df=pd.DataFrame(visitordata)
df = df.resample('A').mean()
df = (df.set_index(df.index)
        .rename_axis("").T)
df.plot()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f7fd62aec90>

(Figure 3: Average view of 12mos/1yr)
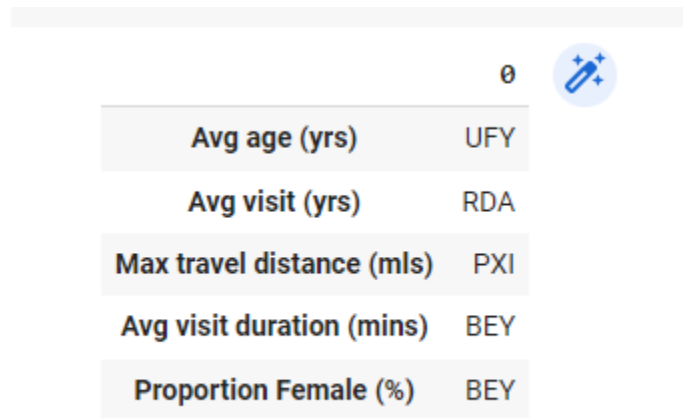
Summary Data

For summary I read all the others CSV. And then stored all into 2nd Dataframe. Then it became 6 rows x 40 columns. (Figure 4)

Id	AEQ	AWF	AXM	BEY	BKI	BQV	CQC	CWN	DKS	GLQ	...	XPE	XXO	YDI	
Avg age (yrs)	32.000000	48.000000	47.000000	39.00000	28.000000	47.000000	23.000000	35.000000	30.000000	23.00000	...	52.00000	54.000000	39.000000	35.000
Avg visit (yrs)	12.361644	139.616438	62.717808	166.29589	29.158904	41.909589	152.578082	211.153425	164.013699	59.20274	...	75.70137	47.342466	27.484932	64.145
Max travel distance (mls)	2.000000	11.000000	8.000000	18.00000	5.000000	5.000000	24.000000	9.000000	21.000000	3.00000	...	4.00000	6.000000	2.000000	6.000
Avg visit duration (mins)	88.000000	105.000000	90.000000	150.00000	101.000000	63.000000	122.000000	90.000000	74.000000	92.00000	...	83.00000	84.000000	64.000000	116.000
Proportion Female (%)	44.000000	54.000000	47.000000	60.00000	42.000000	42.000000	54.000000	59.000000	49.000000	44.00000	...	46.00000	52.000000	42.000000	49.000
Avg spend (£)	12.000000	19.000000	34.000000	18.00000	15.000000	30.000000	9.000000	16.000000	23.000000	16.00000	...	26.00000	38.000000	14.000000	14.000

6 rows × 40 columns

(Figure 4: Gathering all Datas into one DataFrame)

And the datas here are average. For the visit, I have calculated yearly Average (Figure 4). Then I calculated Maximum status for all the Venue ID for those 6 rows (Figure 5).

	0
Avg age (yrs)	UFY
Avg visit (yrs)	RDA
Max travel distance (mls)	PXI
Avg visit duration (mins)	BEY
Proportion Female (%)	BEY

(Figure 5: Maximum status for all Venue ID for each summary.)

We need this because we are going to show the maximum status in chart. Otherwise it will be very messy with much data. So, the output looks like this:
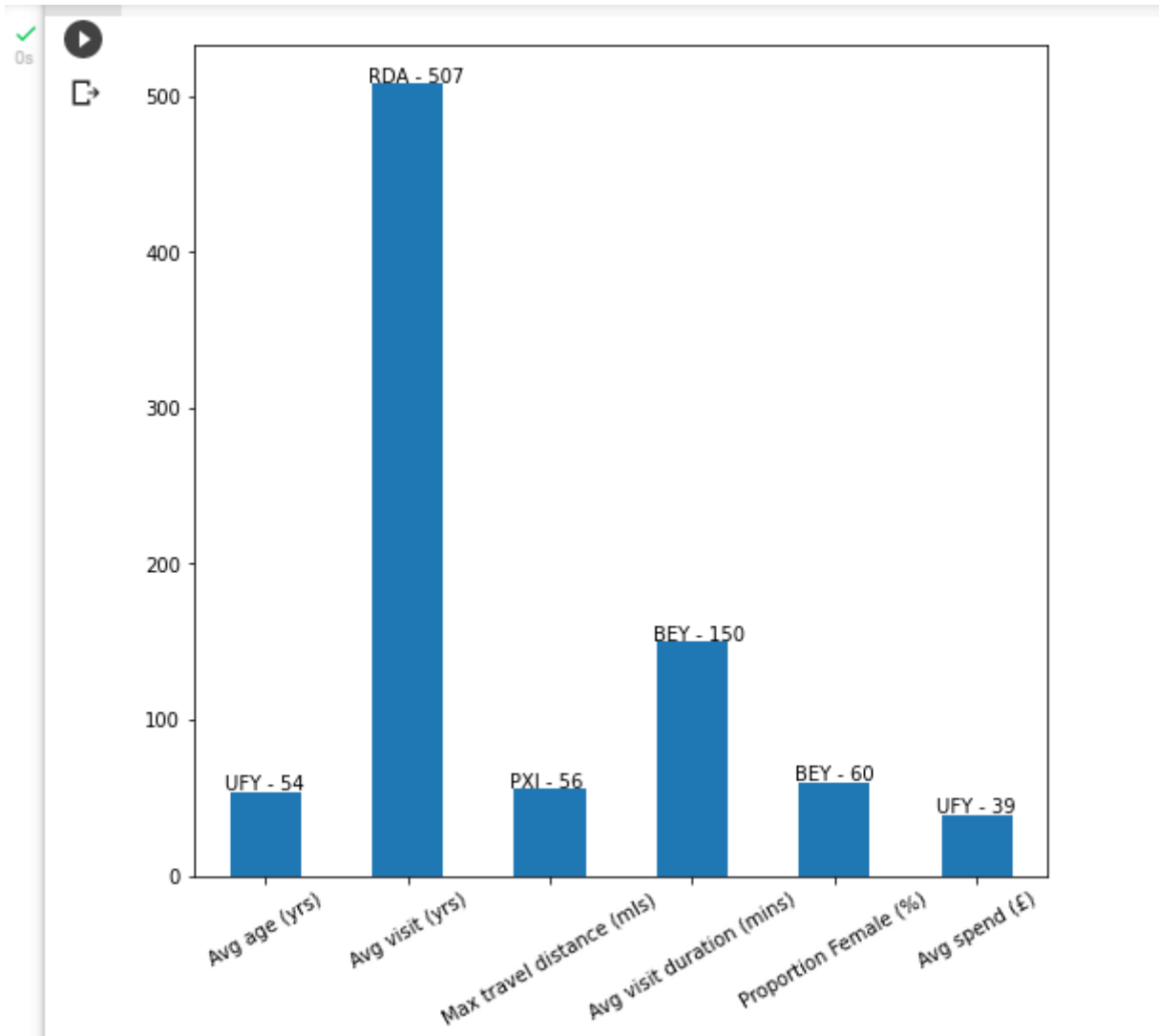
(Figure 6: Maximum status for all the Venue ID for summary in graph)

Critical review

I learnt Pandas, DataFrame, Patch(For setting color), Matplotlib, plot, read_csv, resample, DataFrame mean, indexes, set_indexes, rename_axis, Append, pivot_table, columns, reset_index, iloc, loc, max, Average, to_frame, tolist, head, get_bbox, annotate, xticks etc.

I faced lots of error because of converting Rows to Columns and Converting columns to rows. But at the last I did it successfully.

Data conclusions

The data are of 2019. And the company has 40 Venue.

 ➢ The average age (yrs) is 54 of UFY.
 ➢ The average visit (yrs) is 507 of RDA.

- The max travel distance is 56 miles of PXI.
- The average visit duration is 150 mins (2hrs and 30 mins) of BEY.
- The Proportion Female 60% of BEY.
- The Average spend is 39 euro of UFY.