Data analysis Python project: Money talk

By: Kaede and Kaitlyn

1.
Required tasks

Analyse fake sales data

- Read the data from the spreadsheet

year	month	sales	expenditure
2018	jan	6226	3808
2018	feb	1521	3373
2018	mar	1842	3965
2018	apr	2051	1098
2018	may	1728	3046
2018	jun	2138	2258
2018	jul	7479	2084

DEMO

- Collect all of the sales from each month into a single list

['jan: 6226', 'feb: 1521', 'mar: 1842', 'apr: 2051', 'may: 1728', 'jun: 2138', 'jul: 7479', 'aug: 4434', 'sep: 3615', 'oct: 5472', 'nov: 7224', 'dec: 1812']

- Output the total sales across all months

Total yearly sales: 45542

2.
Extending the project

Analyse fake sales data 2

- Profit per months

['jan: 2418', 'feb: -1852', 'mar: -2123', 'apr: 953', 'may: -1318', 'jun: -120', 'jul: 5395', 'aug: 1635', 'sep: 1966', 'oct: 4356', 'nov: 5793', 'dec: -1720']

- The total expenditure of 2018

Total yearly expenditure: 30159

- The total profit of 2018

Total yearly profit: 15383

- Recreate the csv again

```
year,months,sales,expenditure
2018,jan,6226,3808
2018,feb,1521,3373
2018,mar,1842,3965
2018,apr,2051,1098
2018,may,1728,3046
2018,jun,2138,2258
2018,jul,7479,2084
2018,aug,4434,2799
2018,sep,3615,1649
2018,oct,5472,1116
2018,nov,7224,1431
2018,dec,1812,3532
```

DEMO

3.
Glassdoor analysis

Using pandas and seaborn

Analyse Glassdoor data to see Gender pay gap

- Downloaded a table from kaggle and read in

JobTitle	Gender	Age	PerfEval	Education	Dept	Seniority	BasePay	Bonus
Graphic Designer	Female	18	5	College	Operations	2	42363	9938
Software Engineer	Male	21	5	College	Management	5	108476	11128
Warehouse Associate	Female	19	4	PhD	Administration	5	90208	9268

- Mean pay of all employees

Mean base pay: 94472.653

- Mean pay of male and female

Gender

Female 89942.818376

Male 98457.545113

Name: BasePay, dtype: float64

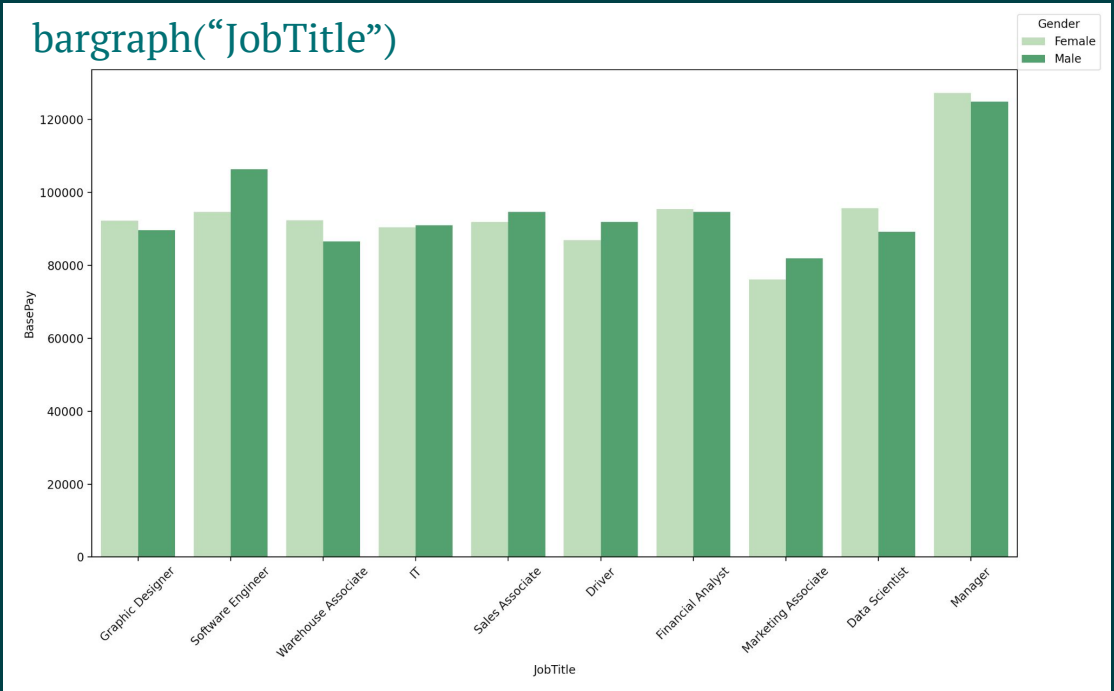
DEMO

Making bar graphs comparing pay between gender

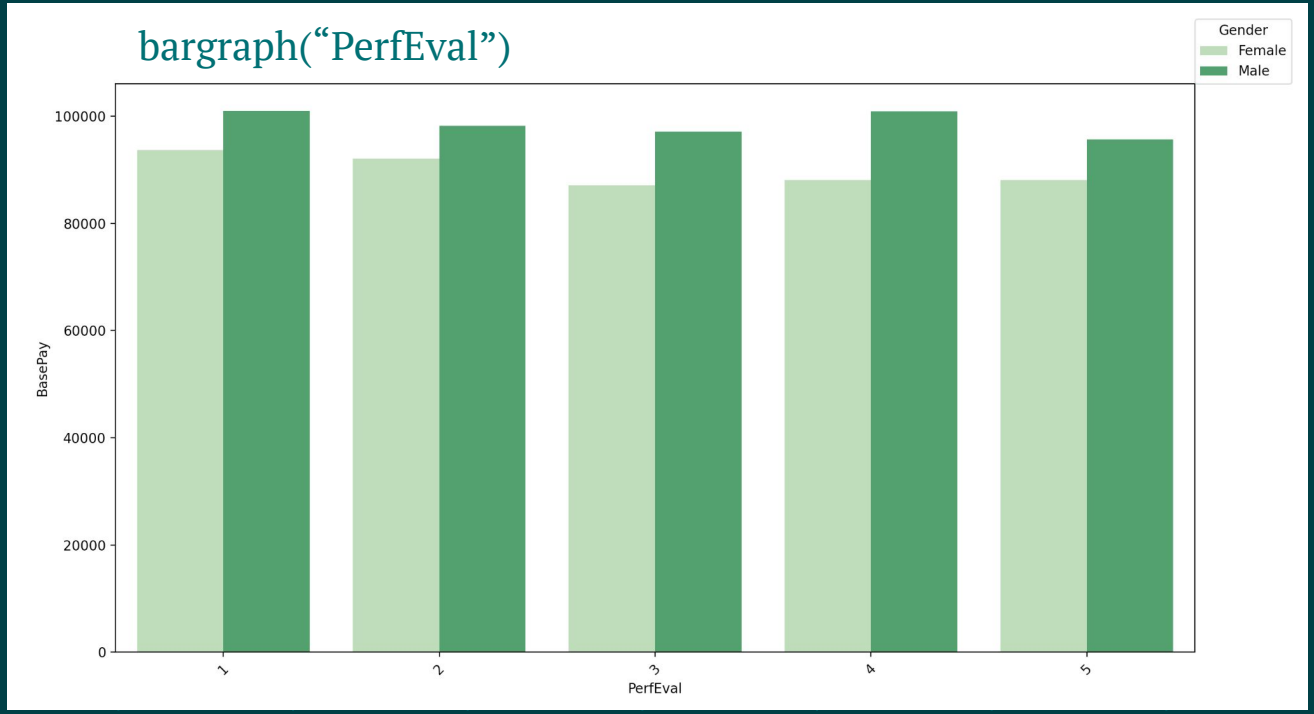
Creating function to make bar graph

```
def bargraph(a):
    plt.figure(figsize=(8, 4))
    sns.barplot(x=a, y='BasePay', hue='Gender', data=df, ci=None,
palette='Greens')
    plt.legend(loc=[1, 1], title='Gender')
    plt.xticks(rotation=45)
    plt.subplots_adjust(bottom=0.2)
    plt.show()
```

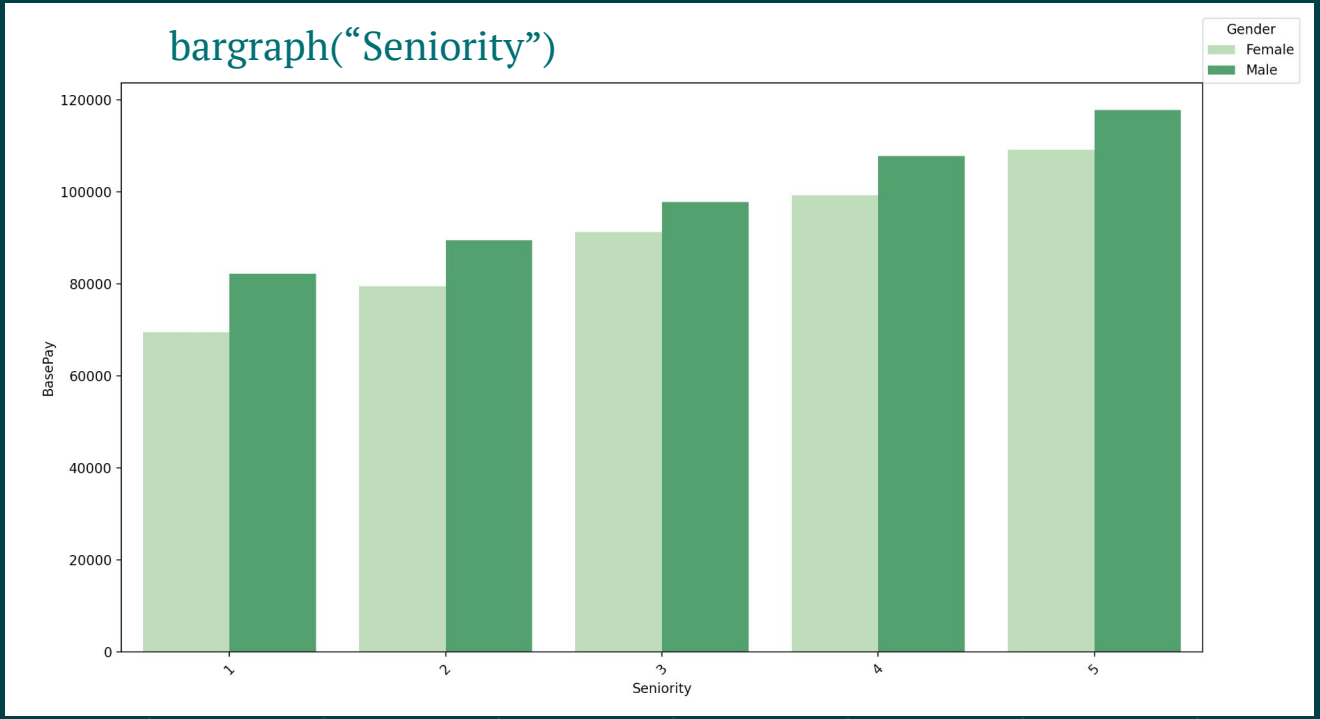
Comparing gender average base pay gap for each job title



Comparing gender average base pay gap for performance evaluation score



Comparing gender average base pay gap for seniority



Thanks!

Any questions?