

Day 11: *Advanced Group By Uses in Pandas*

Taking a break from looking at property prices, Dot went outside to enjoy a classic Emirati breakfast. At a quaint restaurant, they had a delicious plate of shakshuka, or poached eggs with tomatoes, olive oil, onion, and garlic, and a side of khameer bread, a soft, sweet loaf with dates, fennel, saffron, and cardamom. Getting up from their chair with a full belly and a feeling of satisfaction, Dot summoned a taxi and directed the driver to take them over to Dubai Mall, the second-largest mall in the world!

The car pulled up to a massive complex, several stories tall and stretching across for miles, with the look of many different buildings mashed up together, surrounded by a large, glimmering pool of water with a fountain. Inside, Dot looked around in wonder at the several levels of stores that seemingly went on for eternity. It was like an enclosed city in itself, with different neighbourhoods of different types of stores and amenities. Dot journeyed over to the large indoor aquarium and spent most of the day staring happily at the manta rays, sharks, and fish. They successfully avoided spending too much money – after all, they needed to save considering the cost of the trip itself. Dot returned to the hotel with an even warmer feeling about Dubai at the end of a happy day. "I could get used to living here," they said. "Let me look at these property prices again..." Can you help Dot figure out more information about property prices in Dubai?

Tutorial

Within pandas, one of the most essential and useful functions for data analysis is the group by function. During this challenge, we are going to take a look at more complex group by uses.

You might have noticed that the aggregate functions we used with group by yesterday were always applied to all columns. But what if we wanted to apply different functions to different columns? Do we need to do each column individually? If you think there's a more efficient way, you're correct. Pandas has a better way.

```
import pandas as pd
df = pd.read_csv('dubai_properties_data.csv')

# old way - averages of both columns were computed for each neighborhood
df.groupby(['neighborhood'])[['price', 'price_per_sqft']].mean()

# new way - using the aggregate function .agg()
df.groupby(['neighborhood']).agg({'price' : 'mean', 'price_per_sqft' : 'max'})
```

We can use the aggregate function **.agg()** to select the aggregation we want to do for each column. We are using a Python dictionary as a parameter of the **.agg()** function, where the keys of the dictionary are column names, and the values are functions we want to apply. Furthermore, we can even do two aggregations on one column:

```
df.groupby(['neighborhood']).agg({'price' : ['mean', 'max']})
```

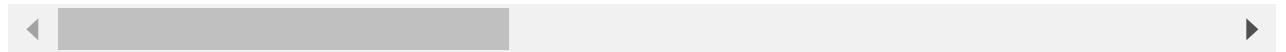
You can visit [the documentation](https://pandas.pydata.org/pandas-docs/stable/10min.html) to have more examples of how to use groupby together with the **.agg()** function.

```
In [1]: import pandas as pd
df = pd.read_csv('dubai_properties_data.csv')
df.head()
```

```
Out[1]:
```

	id	neighborhood	latitude	longitude	price	size_in_sqft	price_per_sqft	no_of_bedrooms
0	5528049	Palm Jumeirah	25.113208	55.138932	2700000	1079	2502.32	1
1	6008529	Palm Jumeirah	25.106809	55.151201	2850000	1582	1801.52	2
2	6034542	Jumeirah Lake Towers	25.063302	55.137728	1150000	1951	589.44	3
3	6326063	Culture Village	25.227295	55.341761	2850000	2020	1410.89	2
4	6356778	Palm Jumeirah	25.114275	55.139764	1729200	507	3410.65	0

5 rows × 38 columns



Challenge

Using the functions described above, which neighborhood has the biggest difference between maximum and minimum property price?

```
In [2]: # SOLUTION

grouped = df.groupby('neighborhood').agg({'price': ['min', 'max']})
grouped.head()
```

```
Out[2]:
```

		price	
		min	max
neighborhood			
Al Barari	1307392	4500000	
Al Barsha	493314	778000	
Al Furjan	350000	1349999	
Al Kifaf	1359000	2100000	
Al Quoz	360000	360000	

```
In [3]: (grouped.iloc[:,1] - grouped.iloc[:,0]).sort_values(ascending=False).head()
```

```
Out[3]:
```

neighborhood	
Palm Jumeirah	34375112
Business Bay	30470000
Jumeirah	26300000
Downtown Dubai	18688888
Jumeirah Beach Residence	14110000

dtype: int64