

An aerial photograph of a dense city skyline, likely New York City, taken from a high vantage point. The image shows a multitude of skyscrapers and buildings, with the Empire State Building and Chrysler Building being prominent. The sky is a deep blue with some light clouds, suggesting dusk or dawn. The overall tone is dark and moody.

Feature Engineering Movie Analytics for Data Scientist



Data Preparation #2

Aggregate

- Aggregate
- Join
- Feature Engineering



Aggregate

- Menghitung nilai rata-rata, median, dan banyak user yang memberikan rating pada film

```
rating_movies = rating.groupby("movieId")["rating"].agg(["mean", "count", "median"])
rating_movies
```

	mean	count	median
--	------	-------	--------

movieId	mean	count	median
---------	------	-------	--------

1	3.872470	247	4.0
---	----------	-----	-----

2	3.401869	107	3.0
---	----------	-----	-----

3	3.161017	59	3.0
---	----------	----	-----

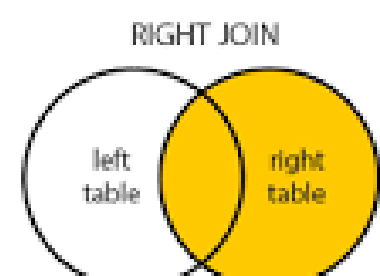
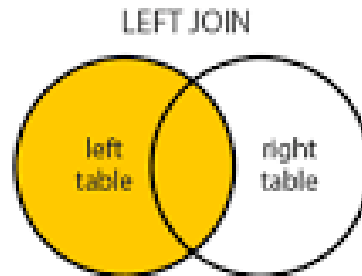
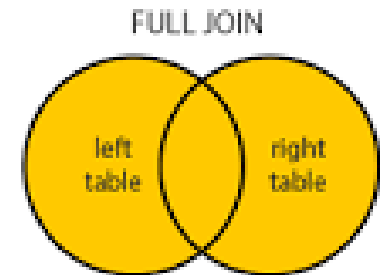
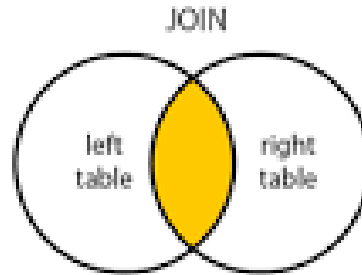
4	2.384615	13	3.0
---	----------	----	-----

5	3.007057	50	3.0
---	----------	----	-----

`.groupby(by = columns, as_index = False)[kolom].agg(stat)`

Merging Data

1. Outer join
2. Inner join
3. Left join
4. Right join
5. Cross Join



Outer Join

Movie ID	Title
1	A
2	B
3	C

Movie ID	Rating
1	5
2	4
4	1

Movie ID	Title	Movie ID	Rating
1	A	1	5
2	B	2	4
3	C		
		4	1

Inner Join

Movie ID	Title
1	A
2	B
3	C

Movie ID	Title	Rating
1	A	5
2	B	4

Movie ID	Rating
1	5
2	4
4	1

Left Join

Movie ID	Title
1	A
2	B
3	C

Movie ID	Title	Rating
1	A	5
2	B	4
3	C	

Movie ID	Rating
1	5
2	4
4	1

Right Join

Movie ID	Title
1	A
2	B
3	C

Movie ID	Title	Rating
1	A	5
2	B	4
4		1

Movie ID	Rating
1	5
2	4
4	1

Cross Join



Movie ID	Title
1	A
2	B
3	C

Movie ID	Rating
1	5
2	4
4	1

Movie ID	Title	Movie ID	Rating
1	A	1	5
1	A	2	4
1	A	4	1
2	B	1	5
2	B	2	4
2	B	4	1
3	C	1	5
3	C	2	4
3	C	4	1

Features Engineering



Operasi Matematika	Operasi Python	Pandas Method
Penjumlahan	+	.add
Pengurangan	-	.sub
Perkalian	*	.mul
Pembagian	/	.div
Pemangkatan	**	.pow

1. Membuat kolom ***profit*** = **Worldwide Box Office – Budget**
2. Membuat kolom Movie yang lebih dari 2 jam
3. Membuat kolom panjang judul film (***title_len***)
4. Membuat kolom runtime dari menit menjadi jam (0-n) (***duraiton_hours***)



A person is captured mid-jump in the center of a city street during the "golden hour" of sunset. The person is wearing a dark jacket and a cap, with their arms outstretched horizontally. The background features a row of bare trees on the left, a bridge, and various city buildings, including a prominent skyscraper with a spire. The sky is a mix of orange, yellow, and blue, with a lens flare effect on the right side. The overall mood is celebratory and energetic.

Thank You!