11 Joins -KirkYagami 👨‍💻🕵️

Splitting the DataFrame

To split the original DataFrame into two DataFrames, you can use various conditions or operations.
Example: Splitting DataFrame Based on Year

```
# Split the DataFrame into two based on year
disney_before_2020 = disney_raw.filter(disney_raw.year < 2020)
disney_2020_and_after = disney_raw.filter(disney_raw.year ⩾ 2020)

# Display the results
disney_before_2020.show(5)
disney_2020_and_after.show(5)
```

Performing Joins

Example: Joining Two DataFrames

Let's say you want to join these DataFrames on a common column, such as `genre` (though in practice, you would join on a column with a common key or ID).

1. **Performing an Inner Join:**

```
# Example of an inner join on the 'genre' column
joined_df = disney_before_2020.join(disney_2020_and_after, on="genre", how="inner")
joined_df.show(5)
```

2. **Performing a Left Outer Join:**

```
# Example of a left outer join on the 'genre' column
joined_df_left = disney_before_2020.join(disney_2020_and_after, on="genre", how="left")
joined_df_left.show(5)
```

3. **Performing a Right Outer Join:**

```
# Example of a right outer join on the 'genre' column
joined_df_right = disney_before_2020.join(disney_2020_and_after, on="genre",
how="right")
joined_df_right.show(5)
```

4. **Performing a Full Outer Join:**

```
# Example of a full outer join on the 'genre' column
joined_df_full = disney_before_2020.join(disney_2020_and_after, on="genre",
```

```
how="outer")
joined_df_full.show(5)
```

Key Points:

- **Join Types:**
  - `inner`: Includes only matching rows from both DataFrames.
  - `left`: Includes all rows from the left DataFrame and matching rows from the right DataFrame.
  - `right`: Includes all rows from the right DataFrame and matching rows from the left DataFrame.
  - `outer`: Includes all rows from both DataFrames, with missing values filled in with nulls where there is no match.
- **Common Columns:** Ensure that the columns you are joining on exist in both DataFrames and are of compatible types.