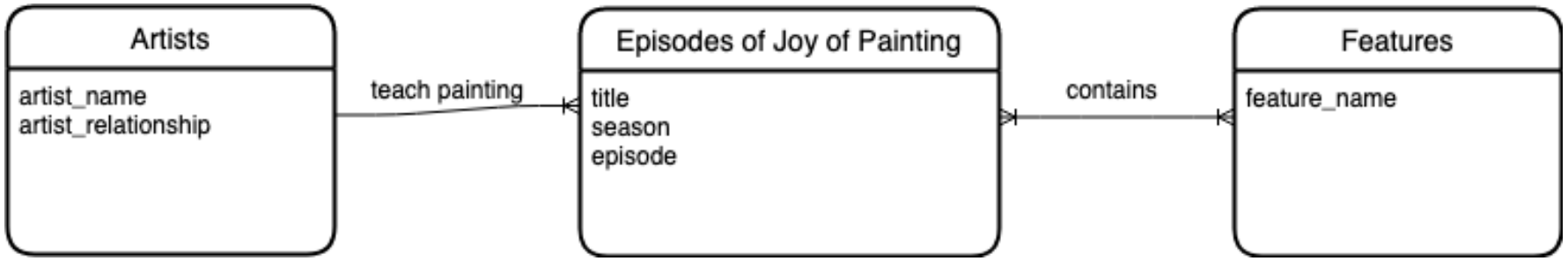
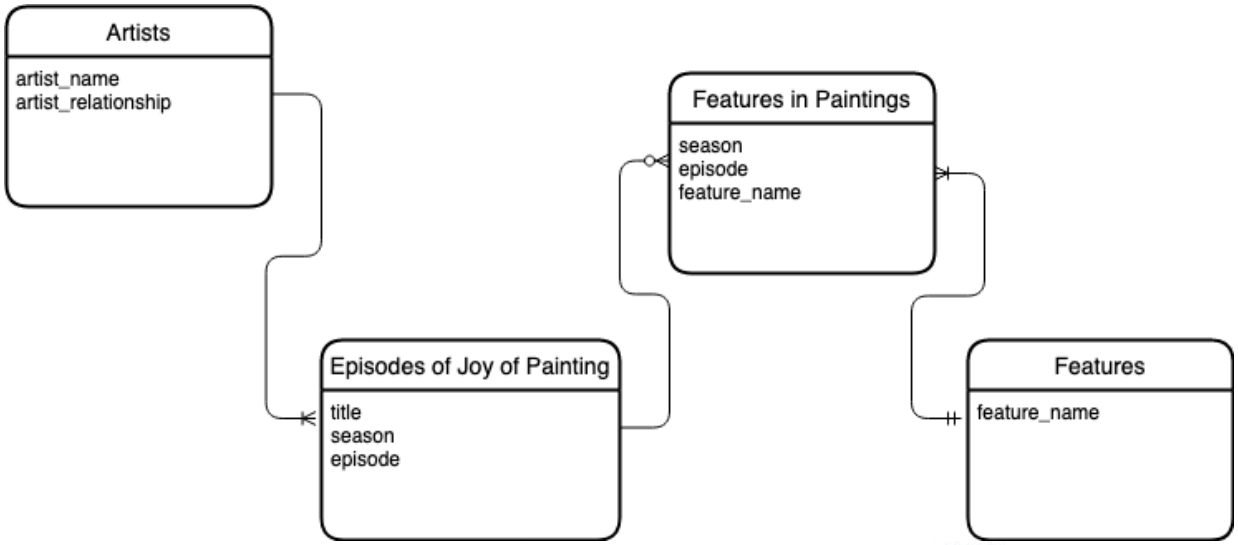
Joy of Painting Database Model

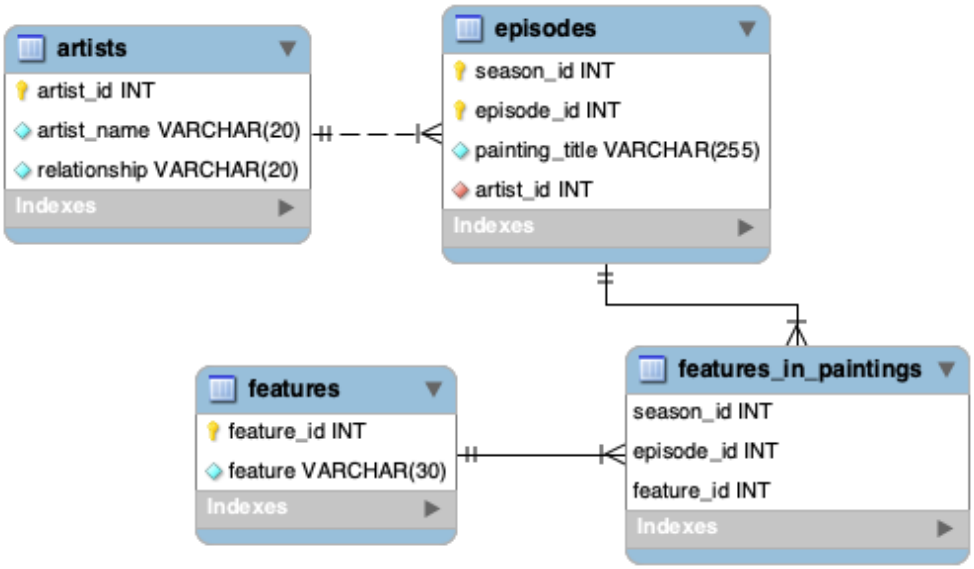
Lisa Over
CMPINF 2110
March 10, 2021



Conceptual Diagram for ross_db



Logical Diagram for ross_db



Physical Diagram for ross_db

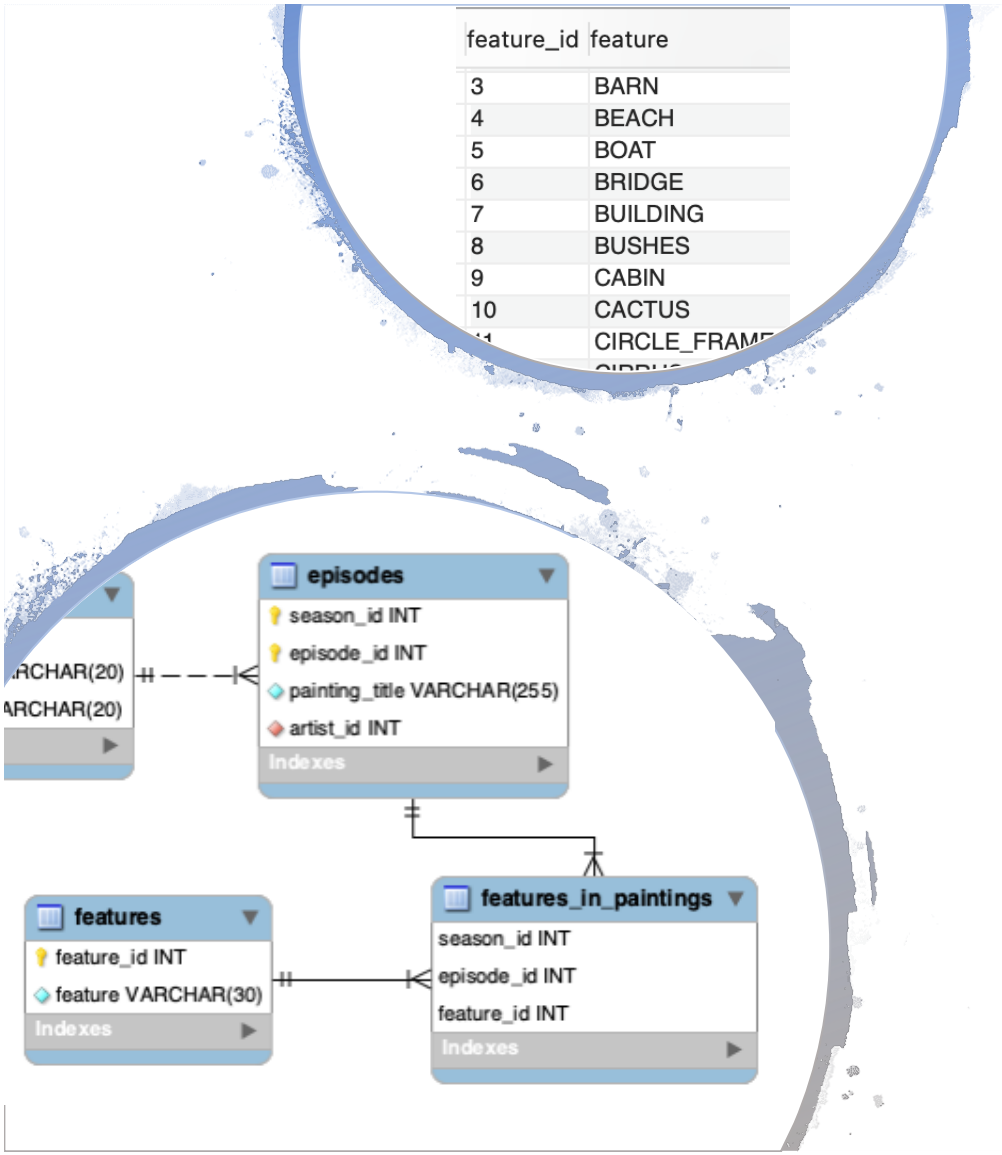
First Normal Form

Tidy data meet 1NF criteria…

- The data is in a two-dimensional table
- There are no repeating groups – no column has more than one value

The data was also in 1NF when the feature names were columns, because each column only had one value. However, that wide format made it difficult to add new or rename features. The wide format also made it more difficult to search the data because one had to include multiple column names in a query rather than searching one column for one or more values in a list.

	season	episode	title	feature	artist_name	artist_relationship
0	1	1	A WALK IN THE WOODS	BUSHES	Bob Ross	Himself
1	1	1	A WALK IN THE WOODS	DECIDUOUS	Bob Ross	Himself
2	1	1	A WALK IN THE WOODS	GRASS	Bob Ross	Himself
3	1	1	A WALK IN THE WOODS	RIVER	Bob Ross	Himself
4	1	1	A WALK IN THE WOODS	TREE	Bob Ross	Himself
...	...	...	...	...	...	...
3185	4	11	NORTHWEST MAJESTY	MOUNTAIN	Diane Andre	Friend
3186	4	11	NORTHWEST MAJESTY	MOUNTAINS	Diane Andre	Friend
3187	4	11	NORTHWEST MAJESTY	SNOWY_MOUNTAIN	Diane Andre	Friend
3188	4	11	NORTHWEST MAJESTY	TREE	Diane Andre	Friend
3189	4	11	NORTHWEST MAJESTY	TREES	Diane Andre	Friend



Second Normal Form

The tidy dataset does not meet 2NF criteria. The nonkey element **feature** is not functionally dependent on the table ID: **season** and **episode**. There are numerous features that could be associated with each key.

To meet 2NF criteria, I moved the features to their own table. Because of the many-to-many relationship between episodes and features, I created the link table *features_in_paintings* to connect *episodes* to *features*. The data now meet 2NF criteria…

- The data still meet 1NF criteria
- The composite key, **season_id** and **episode_id**, determine the **title**, **artist_name**, and **artist_relationship** (there is only one value for each key).

Third Normal Form

The episodes data, even without the features, does not yet meet 3NF criteria. The **artist_relationship** depends on the **artist_name** but is not a candidate key.

To meet 3NF criteria, I moved **artist_name** and **artist_relationship** (as **relationship**) to their own table, *artists*. Because there is a one-to-many relationship between artists and episodes, I made a connection directly from the new *artists* table to the *episodes* table.

The data now meet 3NF criteria because…

- The data still meets 2NF criteria
- There are no transitive dependencies - there are no columns in any table that depend on anything other than their primary key(s)

artist_id	artist_name	relationship
1	Bob Ross	Himself
2	Diane Andre	Friend
3	Other	Friends
4	Steve Ross	Son