Staging

Master_Language (BIRD and data.world): 2790 rows

PK	language	STRING
	country_codes	STRING
	is_official	BOOLEAN
	avg_percentage	FLOAT
	countries	STRING
	total_num_speakers	INTEGER
	degree_of_endangerment	STRING
	data_source	STRING
	load_time	TIMESTAMP
PK	effective_time	TIMESTAMP
	discontinue_time	TIMESTAMP
	status_flag	BOOLEAN

Country_Location (BIRD): 239 rows

PK	code	STRING
	name	STRING
	continent	STRING
	region	STRING
	surface area	FLOAT
	capital	INTEGER
	data_source	STRING
	load_time	TIMSTAMP

City (BIRD_ai): 4079 rows

PK	city_id	INTEGER
	name	STRING
FK	country_code	STRING
	district	STRING
	population	INTEGER
	data_source	STRING
	load_time	TIMESTAMP
	livability	STRING

The livability field was populated using a LLM.

Economic_Political (BIRD): 239 rows

PK, FK	code	STRING
	indep year	INTEGER
	gnp	FLOAT
	gnp old	FLOAT
	government form	STRING
	head of state	STRING
	data_source	STRING
	load_time	TIMESTAMP
	war_probability	STRING

The war_probability field was populated using a LLM.

Social (BIRD): 239 rows

PK, FK	code	STRING
	population	INTEGER
	life expectancy	FLOAT
	local name	STRING
	data_source	STRING
	load_time	TIMESTAMP
	life_expectancy_2100	FLOAT
	life_expectancy_3000	FLOAT
	population_2100	FLOAT
	population_3000	FLOAT

Top_250_Movies (Kaggle): 250 rows

PK	rank	INTEGER
	name	STRING
	year	INTEGER
	rating	FLOAT
	genre	STRING
	certificate	STRING
	run_time	STRING
	tagline	STRING
	budget	STRING
	box_office	STRING
	casts	STRING
	directors	STRING
	writers	STRING
	data_source	STRING
	load_time	TIMESTAMP

Film (Kaggle): 250 rows

PK	rank	INTEGER
	name	STRING
	directors	STRING
FK	d_nationality	STRING
	data_source	STRING
	load_time	TIMESTAMP
	review	STRING

The d_nationality and review fields were populated using a LLM.

The life_expectancy_2100, life_expectancy_3000, population_2100, and poulation_3000 fields were populated using a LLM.