NFL Team Spread & Performance Data Preparation

ELLIOTT NEWMAN, AARON DZABOFF, MITCH LORENZ, JACK MCDEVITT

Problem Proposal

What attributes impact spread predictions, and performances of teams in the NFL?

Initial Dataset

Kaggle

3 main entities

NFL Teams
◦ Team Name, conference, division

NFL Stadiums
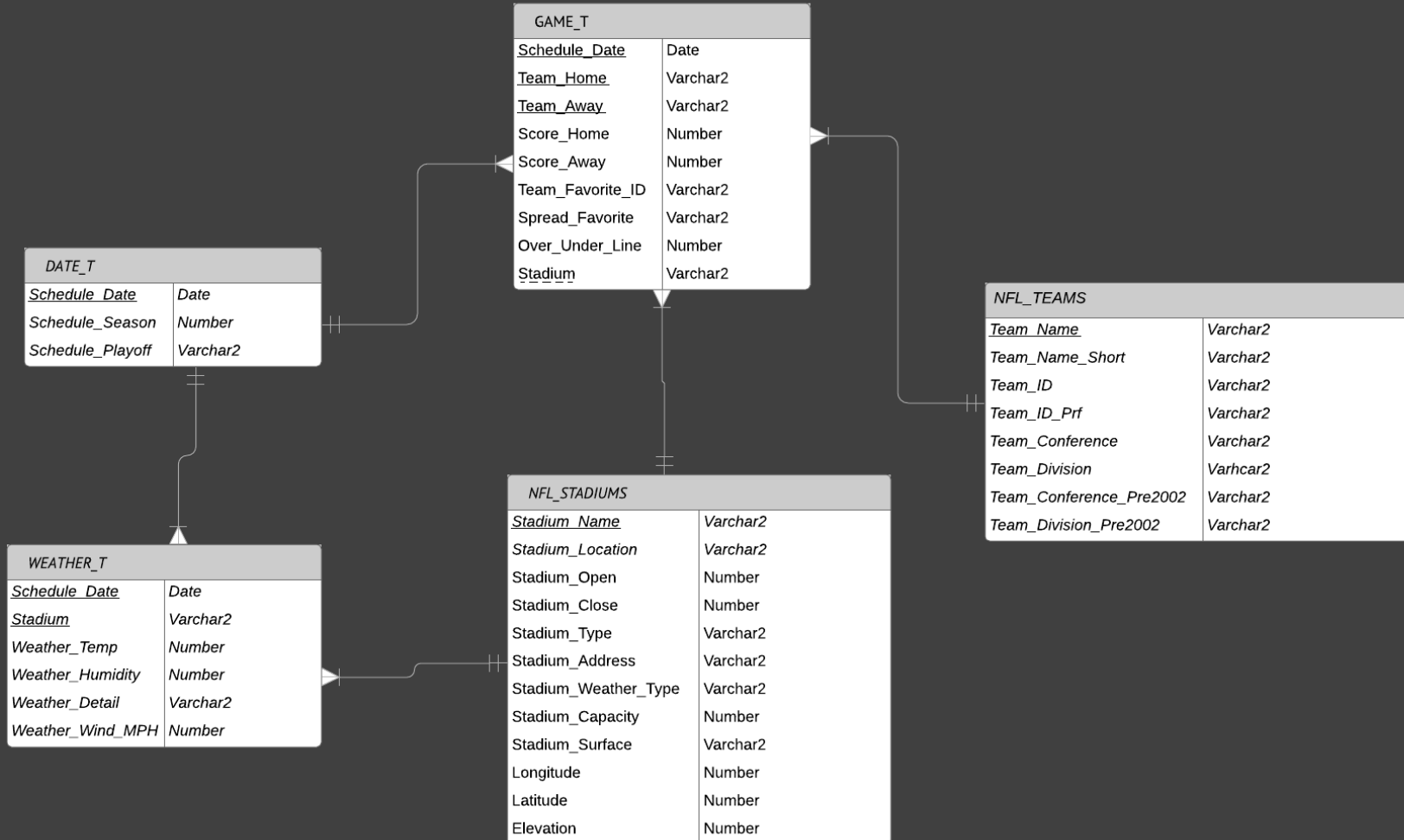◦ Stadium name, location, type, elevation, surface
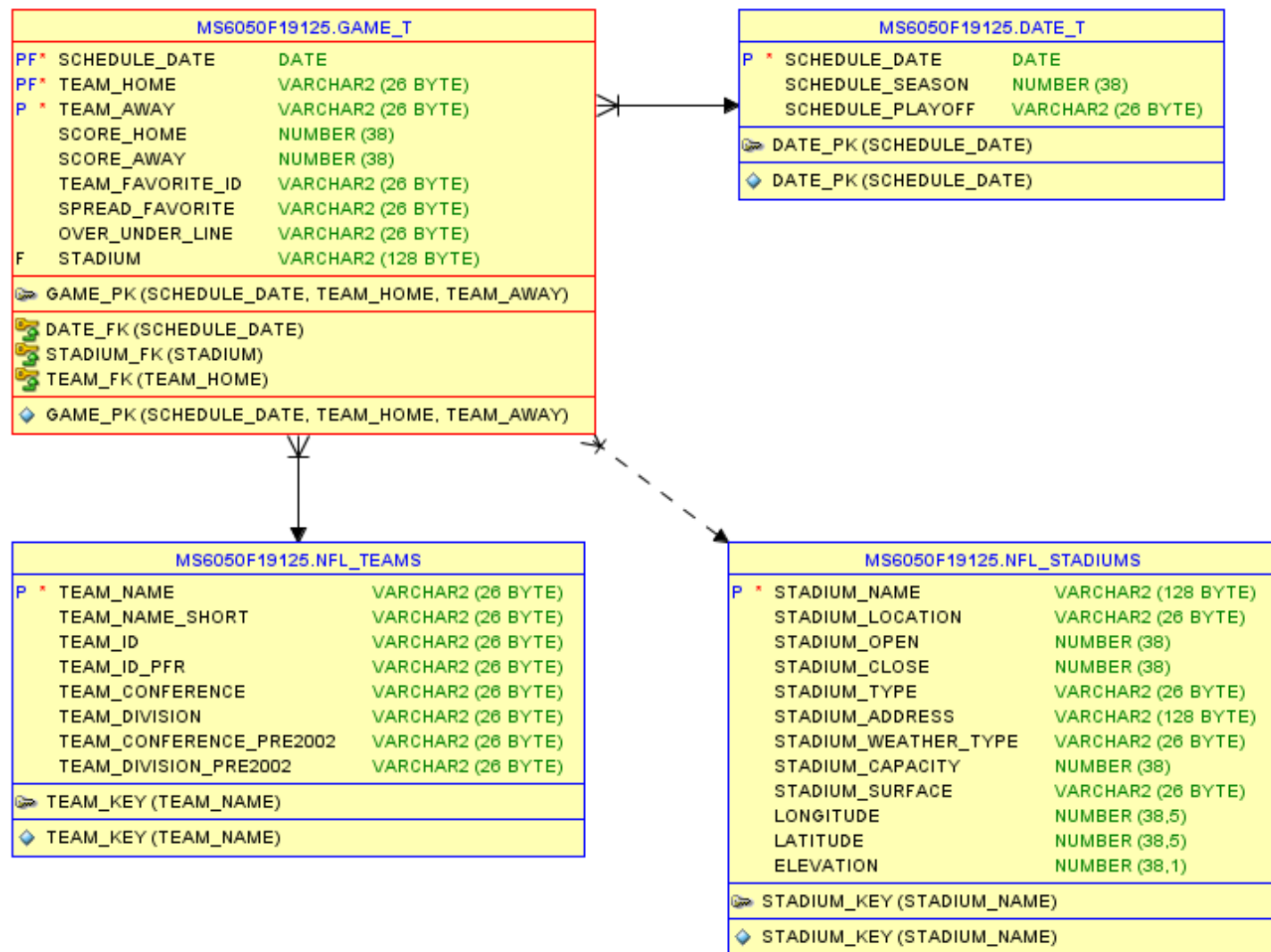
Spreadspoke Gambling Data
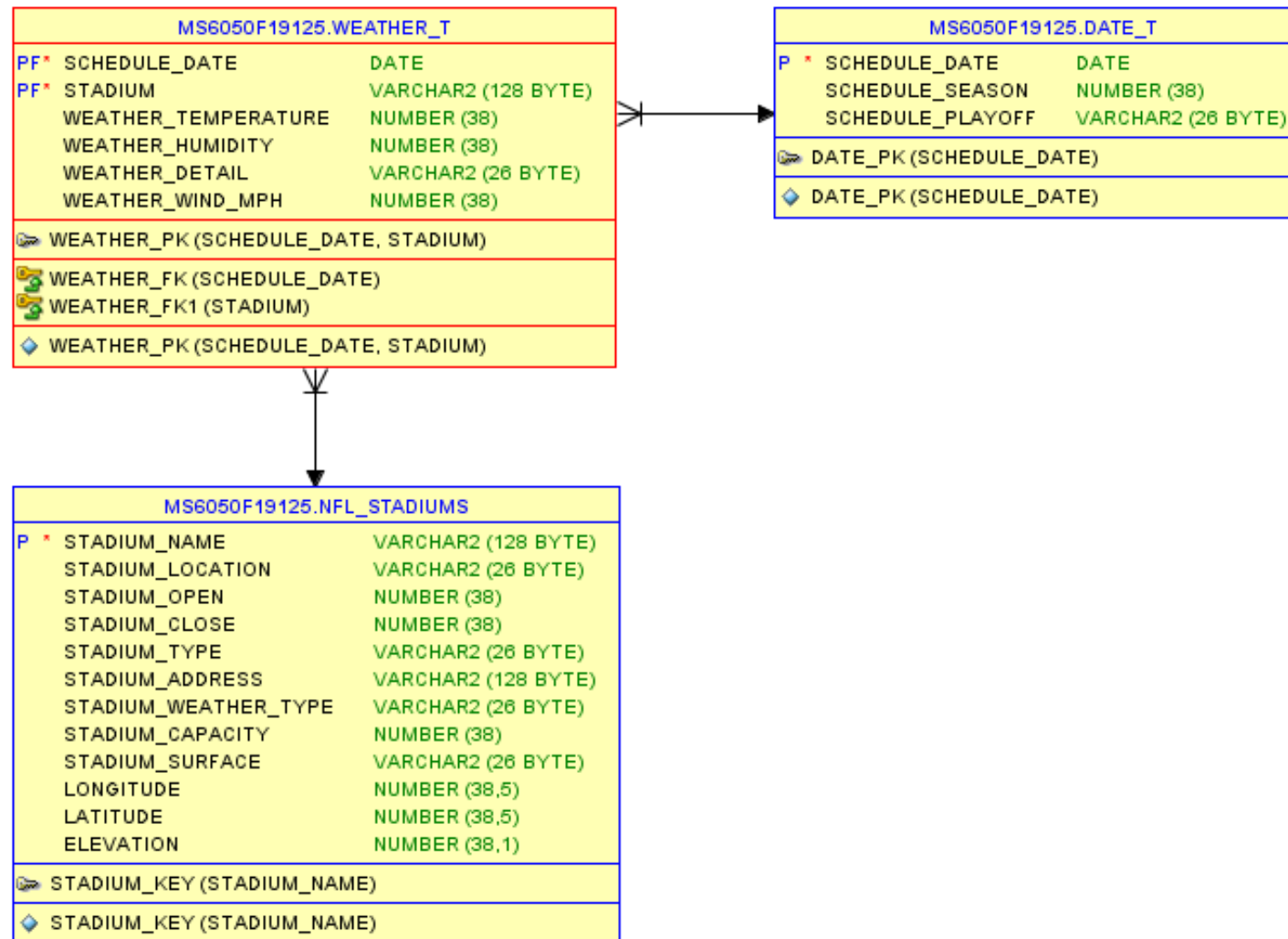◦ Date of Game, home/away team, score, spread, weather

Through normalization, split into 5 main entities

Football Data Mgmt ER Diagram

Elliott Newman | December 1, 2019

GAME_T

Schedule_Date	Date
Team_Home	Varchar2
Team_Away	Varchar2
Score_Home	Number
Score_Away	Number
Team_Favorite_ID	Varchar2
Spread_Favorite	Varchar2
Over_Under_Line	Number
Stadium	Varchar2

DATE_T

Schedule_Date	Date
Schedule_Season	Number
Schedule_Playoff	Varchar2

NFL_TEAMS

Team_Name	Varchar2
Team_Name_Short	Varchar2
Team_ID	Varchar2
Team_ID_Prf	Varchar2
Team_Conference	Varchar2
Team_Division	Varhcar2
Team_Conference_Pre2002	Varchar2
Team_Division_Pre2002	Varchar2

WEATHER_T

Schedule_Date	Date
Stadium	Varchar2
Weather_Temp	Number
Weather_Humidity	Number
Weather_Detail	Varchar2
Weather_Wind_MPH	Number

NFL_STADIUMS

Stadium_Name	Varchar2
Stadium_Location	Varchar2
Stadium_Open	Number
Stadium_Close	Number
Stadium_Type	Varchar2
Stadium_Address	Varchar2
Stadium_Weather_Type	Varchar2
Stadium_Capacity	Number
Stadium_Surface	Varchar2
Longitude	Number
Latitude	Number
Elevation	Number

Data Preparation

Removed invalid data

Removed unnecessary attributes

Updated any incorrect measurements to correct measurements

Created new attributes using SQL 'CASE' statements