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?

Understanding Spreads

CAESARS
ATLANTIC CITY

Initial Dataset

Kaggle

3 main entities

NFL Teams
- Team Name, conference, division

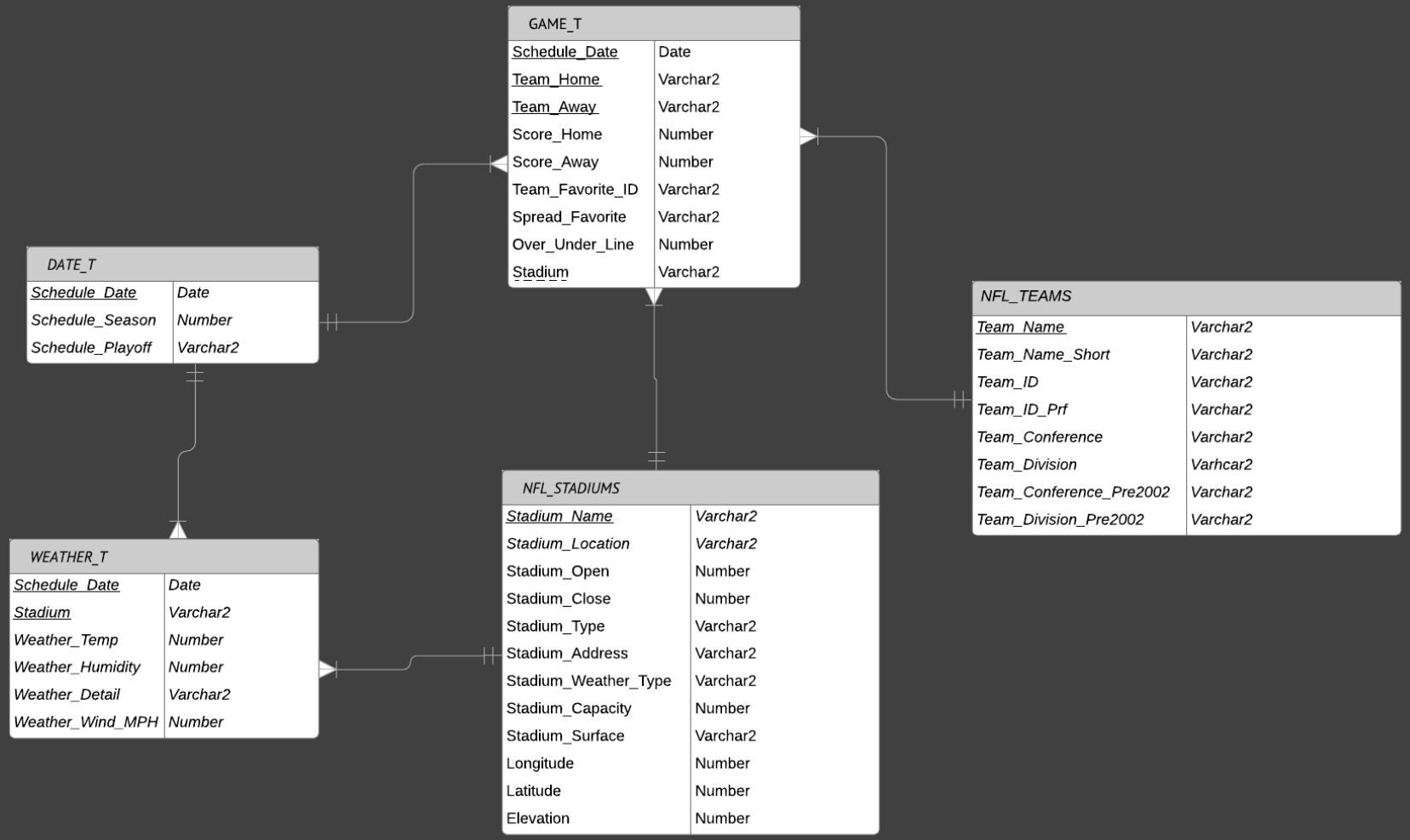
NFL Stadiums
- Stadium name, location, type, elevation, surface

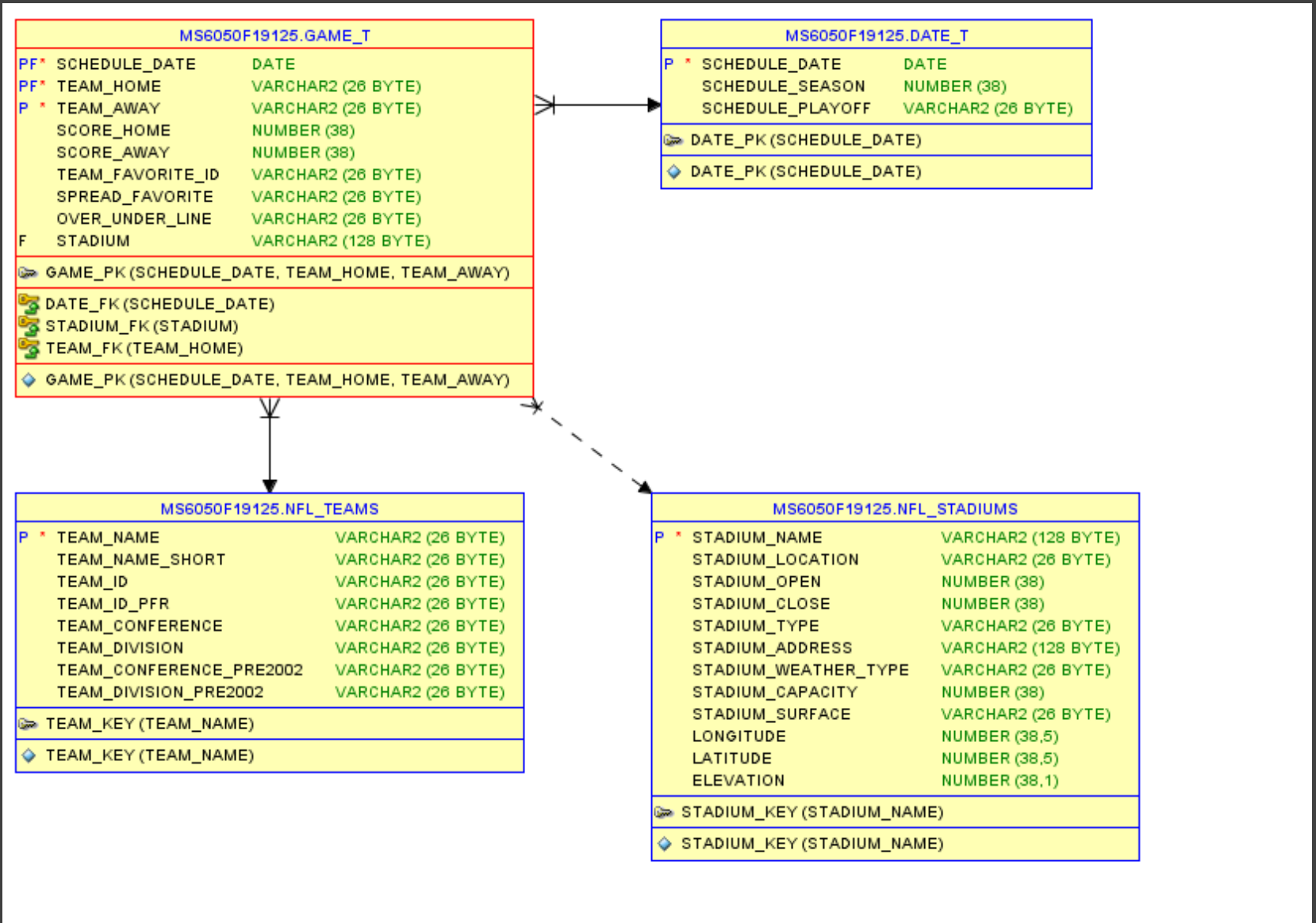
Spreadspoke Gambling Data
- Date of Game, home/away team, score, spread, weather

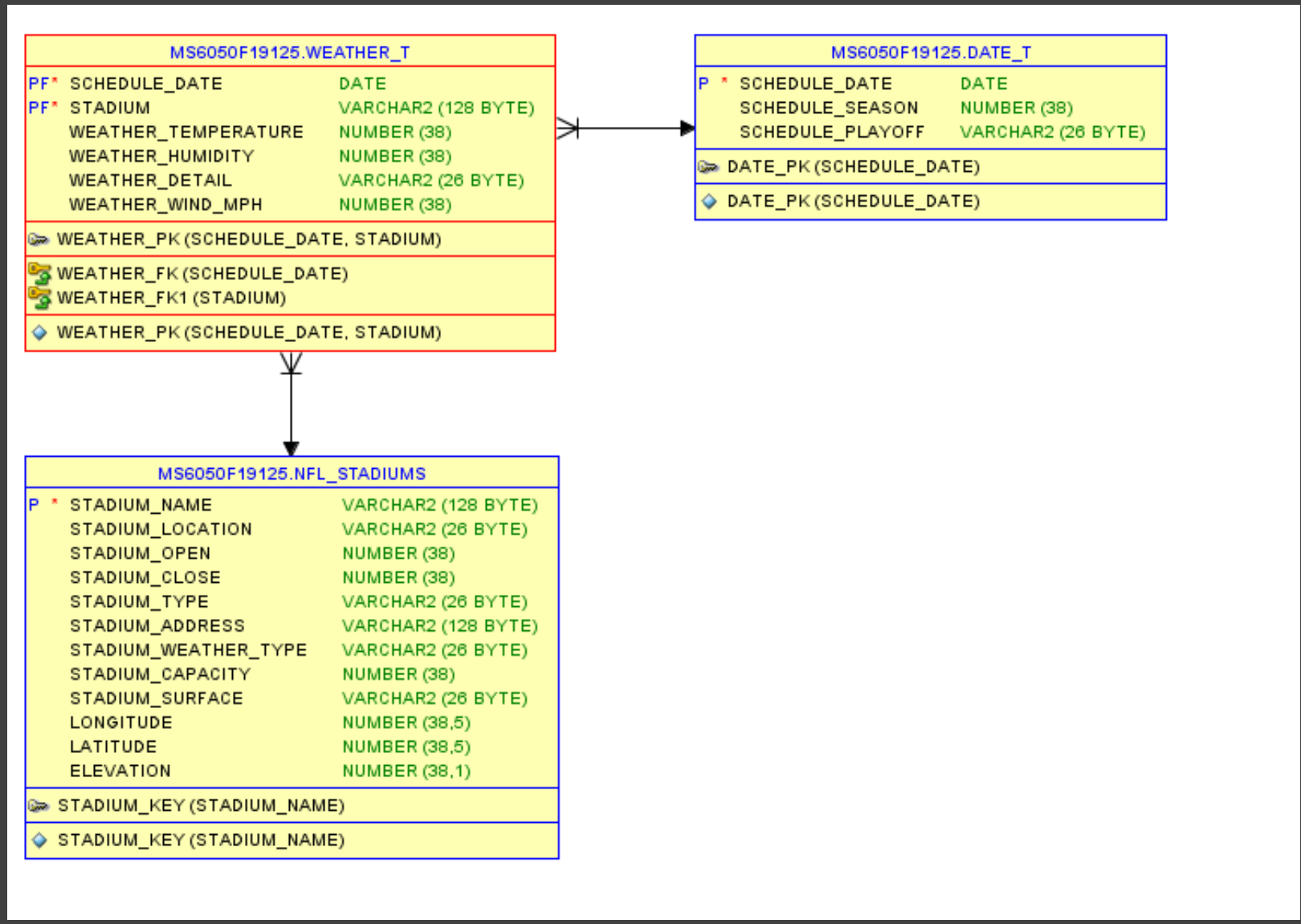
Through normalization, split into 5 main entities

Football Data Mgmt ER Diagram

Elliott Newman | December 1, 2019




MS6050F19125.GAME_T
PF* SCHEDULE_DATE DATE
PF* TEAM_HOME VARCHAR2 (26 BYTE)
P * TEAM_AWAY VARCHAR2 (26 BYTE)
SCORE_HOME NUMBER (38)
SCORE_AWAY NUMBER (38)
TEAM_FAVORITE_ID VARCHAR2 (26 BYTE)
SPREAD_FAVORITE VARCHAR2 (26 BYTE)
OVER_UNDER_LINE VARCHAR2 (26 BYTE)
F STADIUM VARCHAR2 (128 BYTE)
GAME_PK (SCHEDULE_DATE, TEAM_HOME, TEAM_AWAY)
DATE_FK (SCHEDULE_DATE)
STADIUM_FK (STADIUM)
TEAM_FK (TEAM_HOME)
GAME_PK (SCHEDULE_DATE, TEAM_HOME, TEAM_AWAY)
MS6050F19125.DATE_T
P * SCHEDULE_DATE DATE
SCHEDULE_SEASON NUMBER (38)
SCHEDULE_PLAYOFF VARCHAR2 (26 BYTE)
DATE_PK (SCHEDULE_DATE)
DATE_PK (SCHEDULE_DATE)
MS6050F19125.NFL_TEAMS
P * TEAM_NAME VARCHAR2 (26 BYTE)
TEAM_NAME_SHORT VARCHAR2 (26 BYTE)
TEAM_ID VARCHAR2 (26 BYTE)
TEAM_ID_PFR VARCHAR2 (26 BYTE)
TEAM_CONFERENCE VARCHAR2 (26 BYTE)
TEAM_DIVISION VARCHAR2 (26 BYTE)
TEAM_CONFERENCE_PRE2002 VARCHAR2 (26 BYTE)
TEAM_DIVISION_PRE2002 VARCHAR2 (26 BYTE)
TEAM_KEY (TEAM_NAME)
TEAM_KEY (TEAM_NAME)
MS6050F19125.NFL_STADIUMS
P * STADIUM_NAME VARCHAR2 (128 BYTE)
STADIUM_LOCATION VARCHAR2 (26 BYTE)
STADIUM_OPEN NUMBER (38)
STADIUM_CLOSE NUMBER (38)
STADIUM_TYPE VARCHAR2 (26 BYTE)
STADIUM_ADDRESS VARCHAR2 (128 BYTE)
STADIUM_WEATHER_TYPE VARCHAR2 (26 BYTE)
STADIUM_CAPACITY NUMBER (38)
STADIUM_SURFACE VARCHAR2 (26 BYTE)
LONGITUDE NUMBER (38,5)
LATITUDE NUMBER (38,5)
ELEVATION NUMBER (38,1)
STADIUM_KEY (STADIUM_NAME)
STADIUM_KEY (STADIUM_NAME)


MS6050F19125.WEATHER_T
PF* SCHEDULE_DATE DATE
PF* STADIUM VARCHAR2 (128 BYTE)
WEATHER_TEMPERATURE NUMBER (38)
WEATHER_HUMIDITY NUMBER (38)
WEATHER_DETAIL VARCHAR2 (26 BYTE)
WEATHER_WIND_MPH NUMBER (38)
WEATHER_PK (SCHEDULE_DATE, STADIUM)
WEATHER_FK (SCHEDULE_DATE)
WEATHER_FK1 (STADIUM)
WEATHER_PK (SCHEDULE_DATE, STADIUM)
MS6050F19125.DATE_T
P * SCHEDULE_DATE DATE
SCHEDULE_SEASON NUMBER (38)
SCHEDULE_PLAYOFF VARCHAR2 (26 BYTE)
DATE_PK (SCHEDULE_DATE)
DATE_PK (SCHEDULE_DATE)
MS6050F19125.NFL_STADIUMS
P * STADIUM_NAME VARCHAR2 (128 BYTE)
STADIUM_LOCATION VARCHAR2 (26 BYTE)
STADIUM_OPEN NUMBER (38)
STADIUM_CLOSE NUMBER (38)
STADIUM_TYPE VARCHAR2 (26 BYTE)
STADIUM_ADDRESS VARCHAR2 (128 BYTE)
STADIUM_WEATHER_TYPE VARCHAR2 (26 BYTE)
STADIUM_CAPACITY NUMBER (38)
STADIUM_SURFACE VARCHAR2 (26 BYTE)
LONGITUDE NUMBER (38,5)
LATITUDE NUMBER (38,5)
ELEVATION NUMBER (38,1)
STADIUM_KEY (STADIUM_NAME)
STADIUM_KEY (STADIUM_NAME)

Data Preparation

Removed invalid data

Removed unnecessary attributes

Updated any incorrect measurements to correct measurements

Created new attributes using SQL ‘CASE’ statements