IST 659 M408 – Data Admin Concepts & Database Management

2022 MLB Free Agent Tracker

Gianni Conde

Table of Contents

Introduction

Professional sports teams have long been known for their enormous collections of data. With the evolution of technology and growing fan interest, teams have found many uses for their assortments of data, such as player statistics, revenue and transaction catalogs, ballpark concessions, ticket sales, etc. Major League Baseball is considered as the crown jewel of sports data due to their reliance on advanced statistics, known as sabermetrics.

For this project, the dataset that I have chosen is the Fangraphs 2022 MLB Free Agent Tracker. The data set includes 262 free agent eligible players, accompanied by attributes detailing their skillsets and their eventual contracts signed prior to the 2022 season. For this project, we will be observing the top 20 free agents based on WAR for the 2021 season. Our attributes include name, position, bats, throws, previous team, age, service time, 2021 WAR, projected 2022 WAR, QO, signing team, years, total salary, and AAV. The two WAR attributes will be our only statistical identifiers of player evaluation. Our goal is to create a database that discovers trends that could be useful for future free agent predictions that can be used by stakeholders.

This particular database can be invaluable to many stakeholders. Sports agents can use the players listed in this database as benchmarks for their clients to determine the most realistic and negotiable details for their next contract. Fantasy Baseball enthusiasts can use this database to draft skilled players and build successful teams while staying within their allocated parameters. MLB salary arbitrators can use this database to determine the salaries of future arbitration-eligible players by comparing their WAR, projected war, average annual value (AAV), etc.

To ensure the proper management of our data, we will specify a set of business rules. Every attribute in the data set is required except for suffix, qualifying offer (QO), and 2021 WAR. Players who have been traded to a new team during the 2021 season prior to free agency are not eligible to receive a qualifying offer, but since our data does not include multiple previous teams, each player who did not receive one will have an empty space for this attribute. Players who did not play during the 2021 season due to injury or taking their services abroad will not have a value for 2021 WAR. A player must play at least one position. Age, service time, QO, years, total salary, and AAV cannot be negative. 2021 WAR and projected 2022 WAR can be negative. Previous teams and signing teams are not multivalued.

https://www.fangraphs.com/roster-resource/free-agent-tracker

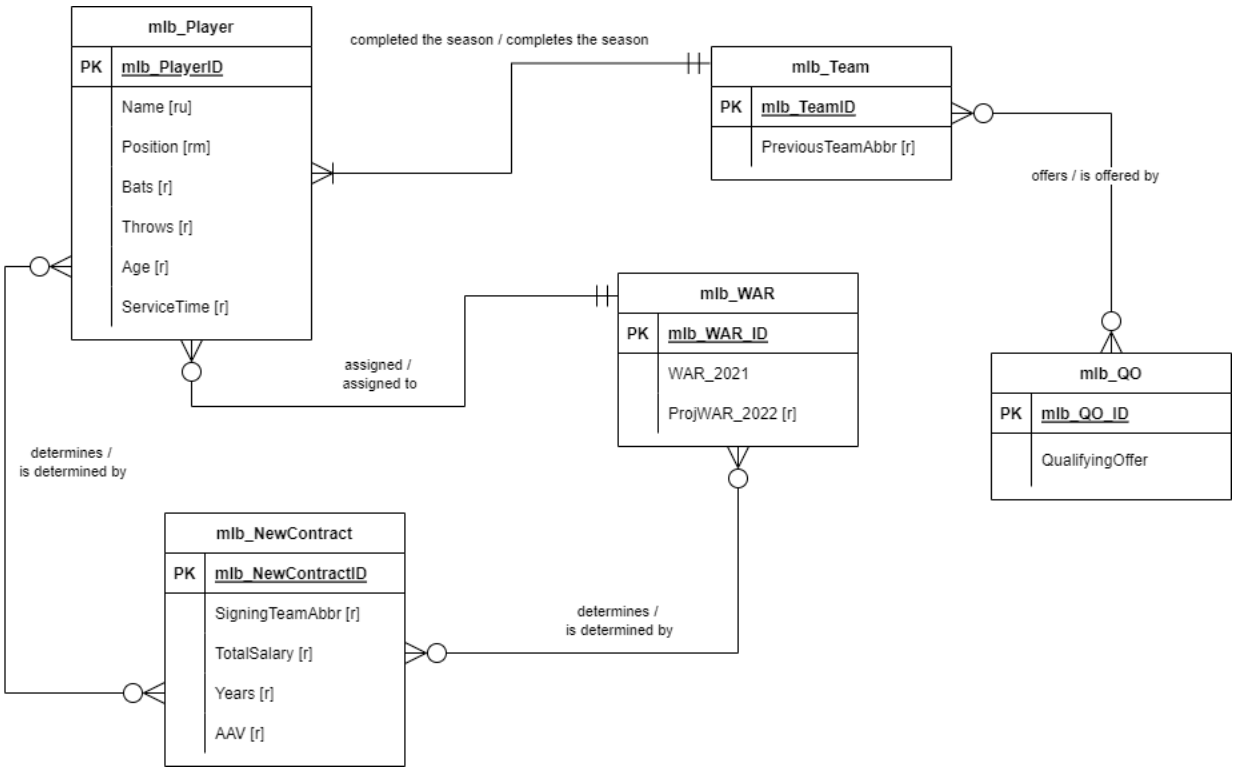
Conceptual Model

Prior to creating our ERD, we will first visualize a list of entities, attributes, and relationships.

Entity	**Attribute**
mlb_Player	Name [ru], Position [rm], Bats [r], Throws [r], Age [r], ServiceTime [r]
mlb_Team	PreviousTeamAbbr [r]
mlb_QO	QualifyingOffer
mlb_WAR	WAR_2021, ProjWAR_2022
mlb_NewContract	SigningTeamAbbr [r], TotalSalary [r], Years [r], AAV [r]

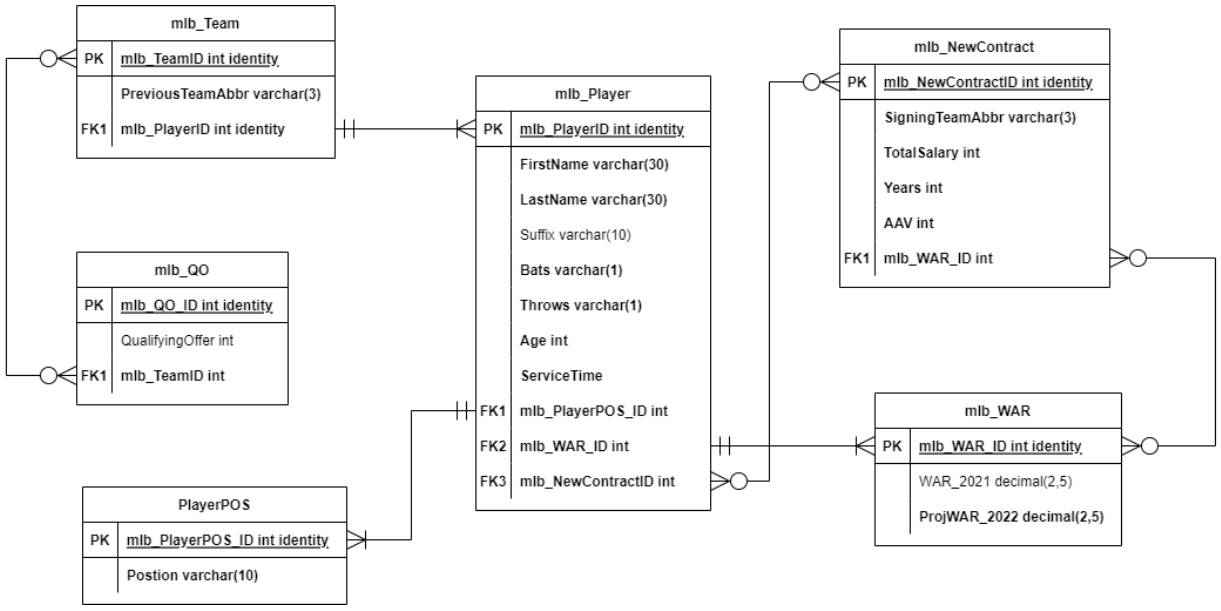
Relationships

- Each free agent eligible player completed the 2021 season with exactly one team; each team completes the season with one or more free agent eligible players.
- Each team offers zero or more qualifying offers; each qualifying offer is offered by zero or more teams.
- Each player is assigned exactly one value of WAR per season; each value of WAR is assigned to all players.
- Each player's mlb_WAR determines their mlb_NewContract's total salary; each mlb_NewContract's total salary is determined by a player's mlb_WAR.
- Each player's age determines their mlb_NewContract's years; each mlb_NewContract's years is determined by a player's age.

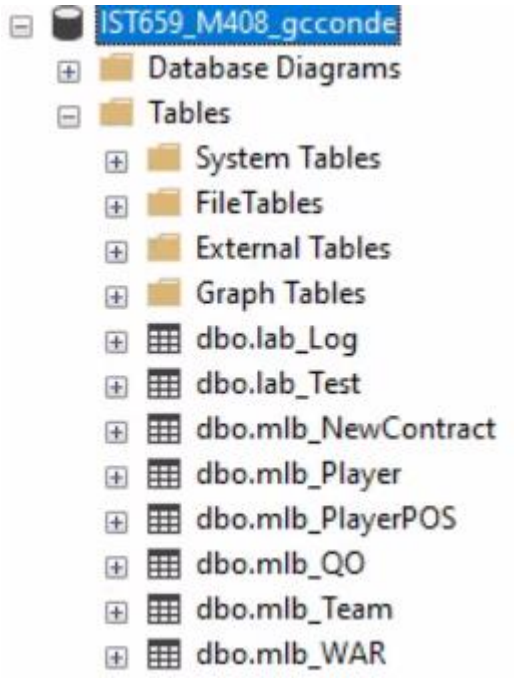


Logical Model

While creating our logical model, we will also be normalizing it so that it will be in third nominal form (3NF). To be in 3NF, a model must first be in 1NF, then 2NF. Since our conceptual model has a multivalued attribute (Position [rm]), it is currently in 0NF. To correct this, we created a new table titled PlayerPosition and transposed the multivalued attribute Position, thus converting the model to 1NF. To convert the model to 2NF, we identified the candidate keys as Position, 2021_WAR, and TotalSalary and removed all PFDs by creating new relations in their place. Lastly, to convert the model to 3NF, we resolved all TFDs by creating new relations with add FKs.



Database Design



SQL Code:

```
-- Creating database tables
-- Table 1: PlayerPOS
CREATE TABLE mlb_PlayerPOS (
        -- Columns for PlayerPOS table
        mlb_PlayerPOS_ID int identity,
        Position varchar(10) NOT NULL,
        -- Constraints for PlayerPOS table
        CONSTRAINT PK_mlb_PlayerPOS PRIMARY KEY (mlb_PlayerPOS_ID)
)
-- End creating PlayerPOS table

-- Table 2: WAR
CREATE TABLE mlb_WAR (
        -- Columns for WAR table
        mlb_WAR_ID int identity,
        WAR_2021 float,
        ProjWAR_2022 float NOT NULL,
        -- Constraints for WAR table
        CONSTRAINT PK_mlb_WAR PRIMARY KEY (mlb_WAR_ID)
)
-- End creating WAR table

-- Table 3: NewContract
```

```
CREATE TABLE mlb_NewContract (
	-- Columns for NewContract table
	mlb_NewContractID int identity,
	SigningTeamAbbr varchar(3) NOT NULL,
	TotalSalary varchar(50) NOT NULL,
	Years int NOT NULL,
	AAV varchar(50) NOT NULL,
	mlb_WAR_ID int NOT NULL,
	-- Constraints for NewContract table
	CONSTRAINT PK_mlb_NewContract PRIMARY KEY (mlb_NewContractID),
	CONSTRAINT FK1_mlb_NewContract FOREIGN KEY (mlb_WAR_ID)
REFERENCES mlb_WAR(mlb_WAR_ID)
)
-- End creating NewContract table

-- Table 4: Player
CREATE TABLE mlb_Player (
	-- Columns for Player table
	mlb_PlayerID int identity,
	FirstName varchar(30) NOT NULL,
	LastName varchar (30) NOT NULL,
	Suffix varchar(10),
	Bats varchar(1) NOT NULL,
	Throws varchar(1) NOT NULL,
	Age int NOT NULL,
	ServiceTime float NOT NULL,
	mlb_PlayerPOS_ID int,
	mlb_WAR_ID int,
	mlb_NewContractID int,
	-- Constraints for Player table
	CONSTRAINT PK_mlb_Player PRIMARY KEY (mlb_PlayerID),
	CONSTRAINT FK1_mlb_Player FOREIGN KEY (mlb_PlayerPOS_ID) REFERENCES
mlb_Player(mlb_PlayerID),
	CONSTRAINT FK2_mlb_Player FOREIGN KEY (mlb_WAR_ID) REFERENCES
mlb_WAR(mlb_WAR_ID),
	CONSTRAINT FK3_mlb_Player FOREIGN KEY (mlb_NewContractID)
REFERENCES mlb_NewContract(mlb_NewContractID)
)
-- End creating Player table

-- Table 5: Team
CREATE TABLE mlb_Team (
	-- Columns for Team table
```

```
	mlb_TeamID int identity,
	PreviousTeamAbbr varchar(3) NOT NULL,
	mlb_PlayerID int NOT NULL,
	-- Constraints for Team table
	CONSTRAINT PK_mlb_Team PRIMARY KEY (mlb_TeamID),
	CONSTRAINT FK1_mlb_Team FOREIGN KEY (mlb_PlayerID) REFERENCES
mlb_Player(mlb_PlayerID)
)
-- End creating Team table

-- Table 6: QO
CREATE TABLE mlb_QO (
	-- Columns for QO table
	mlb_QO_ID int identity,
	QualifyingOffer int,
	mlb_TeamID int,
	-- Constraints for QO table
	CONSTRAINT PK_mlb_QO PRIMARY KEY (mlb_QO_ID),
	CONSTRAINT FK1_mlb_QO FOREIGN KEY (mlb_TeamID) REFERENCES
mlb_Team(mlb_TeamID)
)
-- End creating QO table
-- End creating tables
```

Data Creation

SQL Code:

```
-- Adding data
-- Adding data to the PlayerPOS table
INSERT INTO mlb_PlayerPOS
VALUES
        ('SS'), ('2B/SS'), ('CF/RF/LF'), ('SP'), ('SP'),
        ('1B'), ('SP'), ('SP'), ('SS/2B'), ('SP'), ('RF/LF/DH'),
        ('SS'), ('1B'), ('SP'), ('SP'), ('INF/OF'), ('3B/LF/RF'),
        ('SP'), ('RF/LF/CF'), ('SS')

-- Adding data to the WAR table
INSERT INTO mlb_WAR
VALUES
        ('6.3', '5.3'), ('6.2', '4.5'), ('5.5', '2.8'), ('5.4', '4.2'), ('4.9', '3.6'),
        ('4.8', '4.7'), ('4.8', '3.1'), ('3.9', '3.6'), ('3.9', '2.4'), ('3.8', '3.4'),
        ('3.7', '1.9'), ('3.7', '4.9'), ('3.4', '2.2'), ('3.4', '2.6'), ('3.4', '2.1'),
        ('3.2', '2.1'), ('3.0', '2.3'), ('3.0', '1.4'), ('3.0', '1.9'), ('2.8', '3.8')

-- Adding data to the Player table
INSERT INTO mlb_Player (FirstName, LastName, Suffix, Bats, Throws, Age, ServiceTime)
VALUES
        ('Carlos', 'Correa', NULL, 'R', 'R', '27', '6.119'),
        ('Marcus', 'Semien', NULL, 'R', 'R', '31', '7.118'),
        ('Starling', 'Marte', NULL, 'R', 'R', '33', '8.162'),
        ('Max','Scherzer', NULL, 'R', 'R', '37', '13.079'),
        ('Carlos', 'Rodon', NULL, 'L', 'L', '29', '6.168'),
        ('Freddie', 'Freeman', NULL, 'L', 'R', '32', '11.033'),
        ('Kevin','Gausman', NULL, 'L', 'R', '31', '7.151'),
        (‘Robbie', 'Ray', NULL, 'L', 'L', '30', '7.007'),
        ('Javier', 'Baez', NULL, 'R', 'R', '29', '6.089'),
        ('Eduardo','Rodriguez', 'Jr.', 'L', 'L', '29', '6.130'),
        ('Nick', 'Castellanos', NULL, 'R', 'R', '30', '8.029'),
        ('Corey', 'Seager', NULL, 'L', 'R', '28', '6.032'),
        ('Brandon','Belt', NULL, 'L', 'L', '34', '10.128'),
        ('Clayton', 'Kershaw', NULL, 'L', 'L', '34', '13.105'),
        ('Marcus', 'Stroman', NULL, 'R', 'R', '31', '7.026'),
        ('Chris','Taylor', NULL, 'R', 'R', '31', '6.037'),
        ('Kris', 'Bryant', 'Jr.', 'R', 'R', '30', '6.171'),
        ('Anthony', 'DeSclafani', NULL, 'R', 'R', '32', '7.062'),
        ('Mark','Canha', 'Sr.', 'R', 'R', '33', '6.092'),
        ('Trevor', 'Story', NULL, 'R', 'R', '29', '6.000')
```

```
-- Adding data to the Team table
INSERT INTO mlb_Team (PreviousTeamAbbr, mlb_PlayerID)
VALUES
        ('HOU', '1'), ('TOR', '2'), ('OAK', '3'), ('LAD', '4'), ('CHW', '5'),
        ('ATL', '6'), ('SFG', '7'), ('TOR', '8'), ('NYM', '9'), ('BOS', '10'),
        ('CIN', '11'), ('LAD', '12'), ('SFG', '13'), ('LAD', '14'), ('NYM', '15'),
        ('LAD', '16'), ('SFG', '17'), ('SFG', '18'), ('OAK', '19'), ('COL', '20')

-- Adding data to the QO table
INSERT INTO mlb_QO (QualifyingOffer)
VALUES
        ('18400000'), ('18400000'), (NULL), (NULL),
        (NULL), ('18400000'), (NULL), ('18400000'),
        (NULL), ('18400000'), ('18400000'), ('18400000'),
        ('18400000'), (NULL), (NULL), ('18400000'),
        (NULL), (NULL), (NULL), (NULL), ('18400000')

-- Adding data to NewContract table
INSERT INTO mlb_NewContract (SigningTeamAbbr, TotalSalary, Years, AAV,
mlb_WAR_ID)
VALUES
        ('MIN', '$105,300,000', '3', '$35,100,000', '1'),
        ('TEX', '$175,000,000', '7', '$25,000,000', '2'),
        ('NYM', '$78,000,000', '4', '$19,500,000', '3'),
        ('NYM', '$130,000,000', '3', '$43,333,334', '4'),
        ('SFG', '$44,000,000', '2', '$22,000,000', '5'),
        ('LAD', '$162,000,000', '6', '$24,700,000', '6'),
        ('TOR', '$110,000,000', '5', '$22,000,000', '7'),
        ('SEA', '$115,000,000', '5', '$23,000,000', '8'),
        ('DET', '$140,000,000', '6', '$23,333,334', '9'),
        ('DET', '$77,000,000', '5', '$15,400,000', '10'),
        ('PHI', '$100,000,000', '5', '$20,000,000', '11'),
        ('TEX', '$325,000,000', '10', '$32,500,000', '12'),
        ('SFG', '$18,400,000', '1', '$18,400,000', '13'),
        ('LAD', '$17,000,000', '1', '$17,000,000', '14'),
        ('CHC', '$71,000,000', '3', '$23,666,667', '15'),
        ('LAD', '$60,000,000', '4', '$15,000,000', '16'),
        ('COL', '$182,000,000', '7', '$26,000,000', '17'),
        ('SFG', '$36,000,000', '3', '$12,000,000', '18'),
        ('NYM', '$26,500,000', '2', '$13,250,000', '19'),
        ('BOS', '$140,000,000', '6', '$23,333,334', '20')

-- Verifying added data
```

```
SELECT * FROM mlb_PlayerPOS
SELECT * FROM mlb_WAR
SELECT * FROM mlb_Player
SELECT * FROM mlb_Team
SELECT * FROM mlb_QO
SELECT * FROM mlb_NewContract

/*
	It appears that the wrong datatype was chosen for mlb_QO.QualifyingOffer. Instead of
	QualifyingOffer data being stored as numerical value, we will store it as either being
	‘offered’ or ‘not offered’ with the datatype varchar.
*/
-- End verifying data

-- Updating QO table
ALTER TABLE mlb_QO
ALTER COLUMN QualifyingOffer varchar(50)

TRUNCATE TABLE mlb_QO
SELECT * FROM mlb_QO

-- Adding data to the QO table
INSERT INTO mlb_QO (QualifyingOffer, mlb_TeamID)
VALUES
	('offered', '5'), ('offered', '6'), ('offered', '7'), ('offered', '8'),
	('offered', '9'), ('offered', '10'), ('offered', '11'), ('offered', '12'),
	('offered', '13'), ('offered', '14'), ('offered', '15'), ('offered', '16'),
	('offered', '17'), ('offered', '18'), ('offered', '19'), ('offered', '20'),
	('offered', '21'), ('offered', '22'), ('offered', '23'), ('offered', '24')
-- Verifying QO data
SELECT * FROM mlb_QO
-- End updating QO table
```

Data Questions

1. How many players will have an average ProjWAR_2022?

```
-- Question 1: How many players will have an average ProjWAR_2022?
-- First we will find the average ProjWAR_2022.
SELECT
    AVG(mlb_WAR.ProjWAR_2022) as avg_ProjWAR
FROM mlb_WAR
-- The average ProjWAR_2022 is 3.14.
-- Next, we will discover how many players have a ProjWAR_2022 of average or better.
SELECT
    COUNT(mlb_WAR.ProjWAR_2022) as total_ProjWAR_2022
FROM mlb_WAR
WHERE mlb_WAR.ProjWAR_2022 >= 3.14
-- Lastly, we will list players who have a ProjWAR_2022 >= 3.14.
SELECT
    mlb_Player.FirstName,
    mlb_Player.LastName,
    mlb_WAR.ProjWAR_2022
FROM mlb_Player
JOIN mlb_WAR ON mlb_Player.mlb_WAR_ID=mlb_WAR.mlb_WAR_ID
WHERE mlb_WAR.ProjWAR_2022 >= 3.14
ORDER BY mlb_WAR.ProjWAR_2022 DESC
```

	avg_ProjWAR
1	3.14

	total_ProjWAR_2022
1	9

	FirstName	LastName	ProjWAR_2022
1	Carlos	Correa	5.3
2	Corey	Seager	4.9
3	Freddie	Freeman	4.7
4	Marcus	Semien	4.5
5	Max	Scherzer	4.2
6	Trevor	Story	3.8
7	Robbie	Ray	3.6
8	Carlos	Rodon	3.6
9	Eduardo	Rodriguez	3.4

2. Are older players (32+) guaranteed to have shorter contract lengths (less than 4)?

```
-- Question 2: Are older players (32+) guaranteed to have shorter NewContract lenghts?
-- First we will determine how many players are at least 32 years old.
SELECT mlb_Player.FirstName, mlb_Player.LastName, mlb_Player.Age, mlb_NewContract.Years
FROM mlb_Player
JOIN mlb_NewContract ON mlb_Player.mlb_NewContractID=mlb_NewContract.mlb_NewContractID
WHERE mlb_Player.Age >= 32
ORDER BY mlb_Player.Age DESC
-- It seems that there are 7 players that are 32 years old.
-- Now we will add the 3 year contract maximum to our WHERE statement.
SELECT mlb_Player.FirstName, mlb_Player.LastName, mlb_Player.Age, mlb_NewContract.Years
FROM mlb_Player
JOIN mlb_NewContract ON mlb_Player.mlb_NewContractID=mlb_NewContract.mlb_NewContractID
WHERE mlb_Player.Age >= 32 AND mlb_NewContract.Years < 4
ORDER BY mlb_Player.Age DESC
```

	FirstName	LastName	Age	Years
1	Max	Scherzer	37	3
2	Brandon	Belt	34	1
3	Clayton	Kershaw	34	1
4	Starling	Marte	33	4
5	Mark	Canha	33	2
6	Anthony	DeSclafani	32	3
7	Freddie	Freeman	32	6

	FirstName	LastName	Age	Years
1	Max	Scherzer	37	3
2	Brandon	Belt	34	1
3	Clayton	Kershaw	34	1
4	Mark	Canha	33	2
5	Anthony	DeSclafani	32	3

- Out of the seven players who were 32 years old and older, five of them signed new contracts of less than 4 years. Since contracts less than 4 years are considered to be short-term, we know that older players may not be guaranteed to signed shorter contracts but are more likely to.

3. Are players resigning with their 2021 teams?

```
-- Question 3: Are players resigning with their 2021 teams?
SELECT
    mlb_Player.FirstName,
    mlb_Player.LastName,
    mlb_Team.PreviousTeamAbbr,
    mlb_NewContract.SigningTeamAbbr
FROM mlb_Player
JOIN mlb_Team ON mlb_Player.mlb_PlayerID=mlb_Team.mlb_PlayerID
JOIN mlb_NewContract ON mlb_NewContract.mlb_NewContractID=mlb_Team.mlb_TeamID
WHERE mlb_Team.PreviousTeamAbbr = mlb_NewContract.SigningTeamAbbr
ORDER BY mlb_Team.PreviousTeamAbbr
```

	FirstName	LastName	PreviousTeamAbbr	SigningTeamAbbr
1	Clayton	Kershaw	LAD	LAD
2	Max	Scherzer	LAD	LAD
3	Corey	Seager	LAD	LAD

- By JOINing the mlb_Player, mlb_Team, and mlb_NewContract tables, we can see that there were only three players who resigned with their former teams.

4. Do players with more service time have higher AAV than others?

```
-- Question 4: Do players with more service time have higher AAV than others?
-- ServiceTime in descending order
SELECT mlb_Player.FirstName, mlb_Player.LastName, mlb_Player.ServiceTime, mlb_NewContract.AAV
FROM mlb_Player
JOIN mlb_NewContract ON mlb_Player.mlb_NewContractID=mlb_NewContract.mlb_NewContractID
ORDER BY mlb_Player.ServiceTime DESC
-- AAV in descending order
SELECT mlb_Player.FirstName, mlb_Player.LastName, mlb_Player.ServiceTime, mlb_NewContract.AAV
FROM mlb_Player
JOIN mlb_NewContract ON mlb_Player.mlb_NewContractID=mlb_NewContract.mlb_NewContractID
ORDER BY mlb_NewContract.AAV DESC
```

	FirstName	LastName	ServiceTime	AAV
1	Clayton	Kershaw	13.105	$17,000,000
2	Max	Scherzer	13.079	$43,333,334
3	Freddie	Freeman	11.033	$24,700,000
4	Brandon	Belt	10.128	$18,400,000
5	Starling	Marte	8.162	$19,500,000
6	Nick	Castellanos	8.029	$20,000,000
7	Kevin	Gausman	7.151	$22,000,000
8	Marcus	Semien	7.118	$25,000,000
9	Anthony	DeSclafani	7.062	$12,000,000
10	Marcus	Stroman	7.026	$23,666,667
11	Robbie	Ray	7.007	$23,000,000
12	Kris	Bryant	6.171	$26,000,000
13	Carlos	Rodon	6.168	$22,000,000
14	Eduardo	Rodriguez	6.13	$15,400,000
15	Carlos	Correa	6.119	$35,100,000
16	Mark	Canha	6.092	$13,250,000
17	Javier	Baez	6.089	$23,333,334
18	Chris	Taylor	6.037	$15,000,000
19	Corey	Seager	6.032	$32,500,000
20	Trevor	Story	6	$23,333,334

	FirstName	LastName	ServiceTime	AAV
1	Max	Scherzer	13.079	$43,333,334
2	Carlos	Correa	6.119	$35,100,000
3	Corey	Seager	6.032	$32,500,000
4	Kris	Bryant	6.171	$26,000,000
5	Marcus	Semien	7.118	$25,000,000
6	Freddie	Freeman	11.033	$24,700,000
7	Marcus	Stroman	7.026	$23,666,667
8	Javier	Baez	6.089	$23,333,334
9	Trevor	Story	6	$23,333,334
10	Robbie	Ray	7.007	$23,000,000
11	Kevin	Gausman	7.151	$22,000,000
12	Carlos	Rodon	6.168	$22,000,000
13	Nick	Castella...	8.029	$20,000,000
14	Starling	Marte	8.162	$19,500,000
15	Brandon	Belt	10.128	$18,400,000
16	Clayton	Kershaw	13.105	$17,000,000
17	Eduardo	Rodriguez	6.13	$15,400,000
18	Chris	Taylor	6.037	$15,000,000
19	Mark	Canha	6.092	$13,250,000
20	Anthony	DeSclaf	7.062	$12,000,000

- By JOINing the mlb_Player and mlb_NewContract tables, we can see that Players with more ServiceTime do not necessarily have higher AAVs in their NewContracts. For example, while Max Scherzer has both the highest ServiceTime (13.079) and AAV ($43,333,334), Clayton Kershaw, who has the second highest ServiceTime (13.105), has the 16th highest AAV ($17,000,000).

5. How many Starting Pitchers (SP) received NewContracts of 5 years or more?

```sql
-- Question 5: How many Starting Pitchers (SP) received NewContracts of 5 years or more?
-- First, me must see how many SP are in our database.
SELECT
    mlb_PlayerPOS.Position,
    COUNT(*) as headcount
FROM mlb_PlayerPOS
WHERE mlb_PlayerPOS.Position = 'SP'
GROUP BY mlb_PlayerPOS.Position
ORDER BY headcount DESC
-- We see that there are 8 SPs.
-- Now, we will see how many SPs received NewContracts of 5 years or more.
SELECT
    mlb_Player.FirstName,
    mlb_Player.LastName,
    mlb_PlayerPOS.Position,
    mlb_NewContract.Years
FROM mlb_Player
JOIN mlb_PlayerPOS ON mlb_Player.mlb_PlayerPOS_ID=mlb_PlayerPOS.mlb_PlayerPOS_ID
JOIN mlb_NewContract ON mlb_NewContract.mlb_NewContractID=mlb_PlayerPOS.mlb_PlayerPOS_ID
WHERE mlb_PlayerPOS.Position = 'SP' AND mlb_NewContract.Years >= 5
GROUP BY mlb_Player.FirstName, mlb_Player.LastName, mlb_PlayerPOS.Position, mlb_NewContract.Years
```

	Position	headcount
1	SP	8

	FirstName	LastName	Position	Years
1	Clayton	Kershaw	SP	10
2	Eduardo	Rodriguez	SP	5
3	Max	Scherzer	SP	7
4	Robbie	Ray	SP	6

6. Are players aged 32 and older receiving qualifying offers?

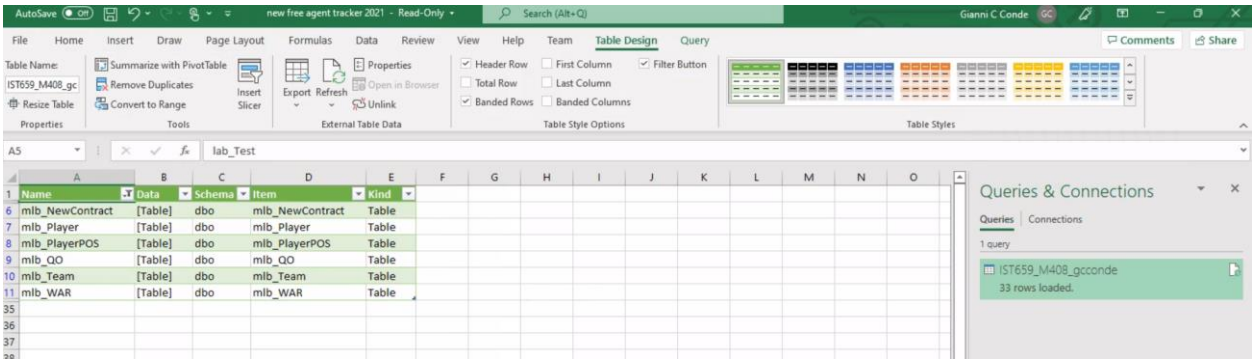
```sql
-- Question 6: Are players aged 32 and older receiving qualifying offers?
-- All players aged 32 and older.
SELECT
    mlb_Player.FirstName,
    mlb_Player.LastName,
    mlb_Player.Age,
    mlb_QO.QualifyingOffer
FROM mlb_Player
JOIN mlb_QO ON mlb_Player.mlb_PlayerID=mlb_QO.mlb_QO_ID
WHERE mlb_Player.Age >= 32
ORDER BY mlb_Player.Age DESC

-- Players 32+ and were offered a QO.
SELECT
    mlb_Player.FirstName,
    mlb_Player.LastName,
    mlb_Player.Age,
    mlb_QO.QualifyingOffer
FROM mlb_Player
JOIN mlb_QO ON mlb_Player.mlb_PlayerID=mlb_QO.mlb_QO_ID
WHERE mlb_Player.Age >= 32 AND mlb_QO.QualifyingOffer = 'offered'
ORDER BY mlb_Player.Age DESC
```

	FirstName	LastName	Age	QualifyingOffer
1	Max	Scherzer	37	not offered
2	Brandon	Belt	34	offered
3	Clayton	Kershaw	34	not offered
4	Starling	Marte	33	not offered
5	Mark	Canha	33	not offered
6	Anthony	DeSclaf...	32	not offered
7	Freddie	Freeman	32	offered

	FirstName	LastName	Age	QualifyingOffer
1	Brandon	Belt	34	offered
2	Freddie	Freeman	32	offered

Implementation

I chose Microsoft Excel as the tool to build a front-end providing use interface for maintenance and query reporting.



	mlb_NewContractID	SigningTeamAbbr	TotalSalary	Years	AAV	mlb_WAR_ID
	3	MIN	$105,300,000	3	$35,100,000	1
	4	TEX	$175,000,000	7	$25,000,000	2
	5	NYM	$78,000,000	4	$19,500,000	3
	6	NYM	$130,000,000	3	$43,333,334	4
	7	SFG	$44,000,000	2	$22,000,000	5
	8	LAD	$162,000,000	6	$24,700,000	6
	9	TOR	$110,000,000	5	$22,000,000	7
	10	SEA	$115,000,000	5	$23,000,000	8
	11	DET	$140,000,000	6	$23,333,334	9
	12	DET	$77,000,000	5	$15,400,000	10
	13	PHI	$100,000,000	5	$20,000,000	11
	14	TEX	$325,000,000	10	$32,500,000	12
	15	SFG	$18,400,000	1	$18,400,000	13
	16	LAD	$17,000,000	1	$17,000,000	14
	17	CHC	$71,000,000	3	$23,666,667	15
	18	LAD	$60,000,000	4	$15,000,000	16
	19	COL	$182,000,000	7	$26,000,000	17

	mlb_PlayerID	FirstName	LastName	Suffix	Bats	Throws	Age	ServiceTime
	1	Carlos	Correa		R	R	27	6.119
	2	Marcus	Semien		R	R	31	7.118
	3	Starling	Marte		R	R	33	8.162
	4	Max	Scherzer		R	R	37	13.079
	5	Carlos	Rodon		L	L	29	6.168
	6	Freddie	Freeman		L	R	32	11.033
	7	Kevin	Gausman		L	R	31	7.151
	8	Robbie	Ray		L	L	30	7.007
	9	Javier	Baez		R	R	29	6.089
	10	Eduardo	Rodriguez	Jr.	L	L	29	6.13
	11	Nick	Castellanos		R	R	30	8.029
	12	Corey	Seager		L	R	28	6.032
	13	Brandon	Belt		L	L	34	10.128
	14	Clayton	Kershaw		L	L	34	13.105
	15	Marcus	Stroman		R	R	31	7.026
	16	Chris	Taylor		R	R	31	6.037
	17	Kris	Bryant	Jr.	R	R	30	6.171

	A	B
1	mlb_PlayerPOS_ID	Position
2	1	SS
3	2	2B/SS
4	3	CF/RF/LF
5	4	SP
6	5	SP
7	6	1B
8	7	SP
9	8	SP
10	9	SS/2B
11	10	SP
12	11	RF/LF/DH
13	12	SS
14	13	1B
15	14	SP
16	15	SP
17	16	INF/OF
18	17	3B/LF/RF

	A	B	C
1	mlb_QO_ID	QualifyingOffer	mlb_TeamID
2	1	offered	5
3	2	offered	6
4	3	not offered	7
5	4	not offered	8
6	5	not offered	9
7	6	offered	10
8	7	not offered	11
9	8	offered	12
10	9	not offered	13
11	10	offered	14
12	11	offered	15
13	12	offered	16
14	13	offered	17
15	14	not offered	18
16	15	not offered	19
17	16	offered	20
18	17	not offered	21

	A	B	C
1	mlb_TeamID	PreviousTeamAbbr	mlb_PlayerID
2	5	HOU	1
3	6	TOR	2
4	7	OAK	3
5	8	LAD	4
6	9	CHW	5
7	10	ATL	6
8	11	SFG	7
9	12	TOR	8
10	13	NYM	9
11	14	BOS	10
12	15	CIN	11
13	16	LAD	12
14	17	SFG	13
15	18	LAD	14
16	19	NYM	15
17	20	LAD	16
18	21	SFG	17

	A	B	C
1	mlb_WAR_ID	WAR_2021	ProjWAR_2022
2	1	6.3	5.3
3	2	6.2	4.5
4	3	5.5	2.8
5	4	5.4	4.2
6	5	4.9	3.6
7	6	4.8	4.7
8	7	4.8	3.1
9	8	3.9	3.6
10	9	3.9	2.4
11	10	3.8	3.4
12	11	3.7	1.9
13	12	3.7	4.9
14	13	3.4	2.2
15	14	3.4	2.6
16	15	3.4	2.1
17	16	3.2	2.1
18	17	3	2.3

Reflection

During the course of this process, some of the things that changed were the entity names, the amount of data to be included, and data types within the entities. Changing the entity name 'Contract' to 'NewContract' relieved some confusion since the phrase 'Contract' was function. Adding 'mlb_' to the beginning of entity names allowed my created tables to be viewed in the Object Explorer together instead of scattered around tables previously created from Lab assignments. The original data set contained 262 players, so minimizing it to the top 20 players based on 2021 WAR allowed the process to become less tedious while providing adequate analysis and maintaining the integrity of the project. Changing the fields that contained values in dollar figure from integers to varchars, such as QualifyingOffer, allowed for a smoother process of analyses since the QualifyingOffer fields provided the same figures for those who received them. This essentially made the data questions involving this field as yes or no questions.

I believe that I rushed the adding of data process, which led to the altering of tables. Correcting these errors took a great amount of time, so I will be sure to take more care during this process the next time around.

If I were to redo this project, I would replace some unnecessary fields such as Suffix, Bats, and Throws with a player's injury history, such as instances where they were placed on the Injured List, the scope of these injuries, and injury descriptions. I believe that these factors contribute to a players future contract since teams would be weary of signing injury-prone players. Injuries also vary depending on a player's position, particularly pitchers and position players. I believe that these factors could lead to more advanced data questions. Some examples include:

- Is a pitcher with a 60 day and over injury stint less likely to receive a Qualifying Offer than a position player with an injury stint of over 60 days?
- Are players over the age of 30 less likely to receive a long-term contract (5+ years) if they were placed on the Injured List 3 times or more?

Glossary

Player	Player who is a free agent
Position	Position(s) played
PlayerPOS	Bridge table indicating a player's position
Bats	Bat handedness
Throws	Throwing hand
Age	Age in years
ServiceTime	Years spent on the major league roster
PreviousTeamAbbr	Abbreviation of the player's team during the 2021 season
QO	Qualifying offer
WAR	Wins above replacement
WAR_2021	Player's WAR from the 2021 season
ProjWar_2022	Player's projected WAR for the 2022 season
NewContract	Player's contract signed beginning in the 2022 season
SigningTeamAbbr	Abbreviation of the team that signs a player for the 2022 season
TotalSalary	Total salary in dollars
Years	Length of the contract in years
AAV	Average annual salary (TotalSalary / Years) in dollars