BSD 2343 DATA WAREHOUSING

2021/2022 SEMESTER II

TITLE:

OLAP-BASED ANALYSIS OF DATA FINDING THE MOST COST EFFECTIVE, UNDERRATED, OR SKILLED BASKETBALL PLAYERS.

PREPARED FOR

DR Nor Azuana Ramli

STUDENT NAME	STUDENT ID
SOON KIEN YUAN	SD20040
TAN WEI QING	SD20030
NG JIE HAO	SD20036
AHMAD AMIRUL AUFA BIN AHMAD SHAKIR	SD20005
MUHAMMAD HAZIQ BIN ZAIRUL EFFENDI	SD20026

TABLE OF CONTENTS

1.0 Background

1.1 Description of the project

Basketball is a team sport well-known and supported by many people throughout the whole world. The sport is full of competition, with players from a variety of backgrounds and skills. In fact, according to stats from Sports Show, basketball is earth's third most popular sport with roughly 2.2 billion fans. One quote said, “The strength of the team is each individual member. The strength of each member is the team.”. As a new team, the problem of competing in higher-level tournaments and leagues requires talented players on the team roster. Also, the difference in the player’s market value lies in their skill and performance in the game. Skilled players have higher pay while the less skilled ones have lower pay. The players also have many criteria needed to fulfil, unpreferable body proportions can heavily influence a player’s performance in a game and their dexterity and agility is also a needed talent in a sports game like basketball. This project is aimed to help in properly choosing a roster that is the most optimal in terms of player value and performance with a minimum budget.

1.2 Problem to be solved

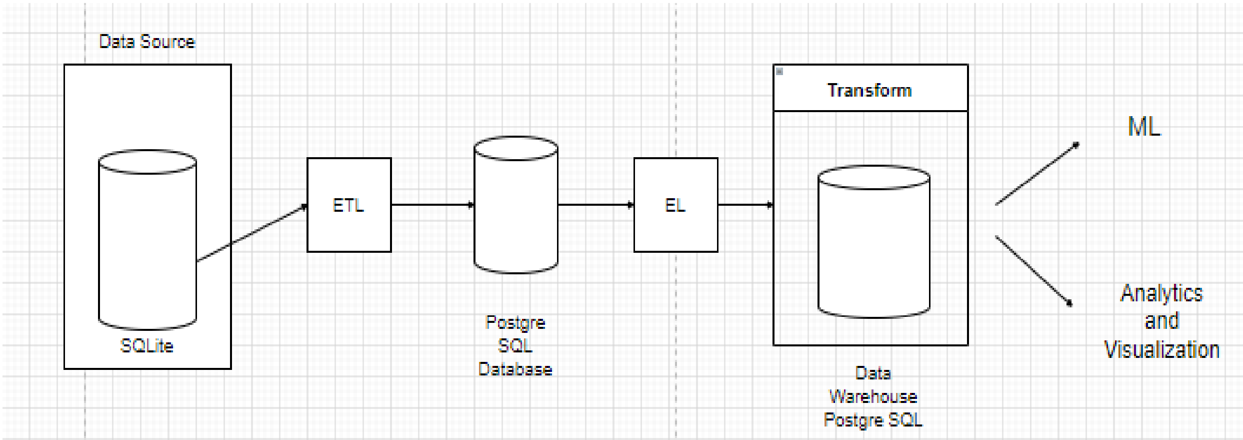
The main issue in building a team is the limited budget presented to a team that is only newly formed. Besides, only a limited scope of players is available to be recruited. Many more issues arise from this as well, one of them being a mediocre performance achieved by the players in international tournaments where major teams compete in. With a small budget, there are chances of unskilled players being included in the roster. Therefore, the solution is to find players that have adequate skills and pay but can further show their potential on the field by playing the correct position in the games.

1.3 Data schema

Figure below shown all table under public schema of the database or basket_database. For the column or entities, it will be available in ERD through a Google Drive link in Appendix 1.

Type	Name
Table	public.Draft
Table	public.Draft_Combine
Table	public.Game
Table	public.Game_Inactive_Players
Table	public.Game_Officials
Table	public.News
Table	public.News_Missing
Table	public.Player
Table	public.Player_Attributes
Table	public.Player_Bios
Table	public.Player_Photos
Table	public.Player_Salary
Table	public.Team
Table	public.Team_Attributes
Table	public.Team_History
Table	public.Team_Salary

2.0 Architecture



Before we go into the details of the flow of the process, we would first like to discuss what data sources we retrieved from. SQLite is a tiny, quick, self-contained, high-reliability, full-featured SQL embedded database engine that written in C language.

Data sources will be extracted and transformed into a specified format that can be accepted by the relational database management system (RDBMS), PostgreSQL database. After the format is transformed, the data source is loaded into the database. Therefore, the database will be the platform for online transaction processing (OLTP), where online transactions like INSERT, UPDATE, DELETE occur.

According to Founder of the Italian PostgreSQL Gabriele Bartolini, PostgreSQL is an ideal candidate of storage system for data warehouse and RDBMS. It means PostgreSQL will be our data warehouse too. Since we use same platform for data warehouse and RDBMS, then extract, load, and transform (ELT) will be the best solution to improve the efficiency of a process and save time. However, there are some data in RDBMS are not suitable in data warehouse for Online analytical processing (OLAP). Hence, transforming step occur in the data warehouse. We will extract the data from the RDBMS and load into database and then transformed the data. Take note that ELT process is supported in PostgreSQL according to Sebastian Insausti, he is support engineer in Severalnines.

Lastly, Machine learning and analytic included OLAP and visualization can be performed by connecting to the data warehouse. In detail, Tableau will used to perform the OLAP and visualization by connecting to the data warehouse we developed.

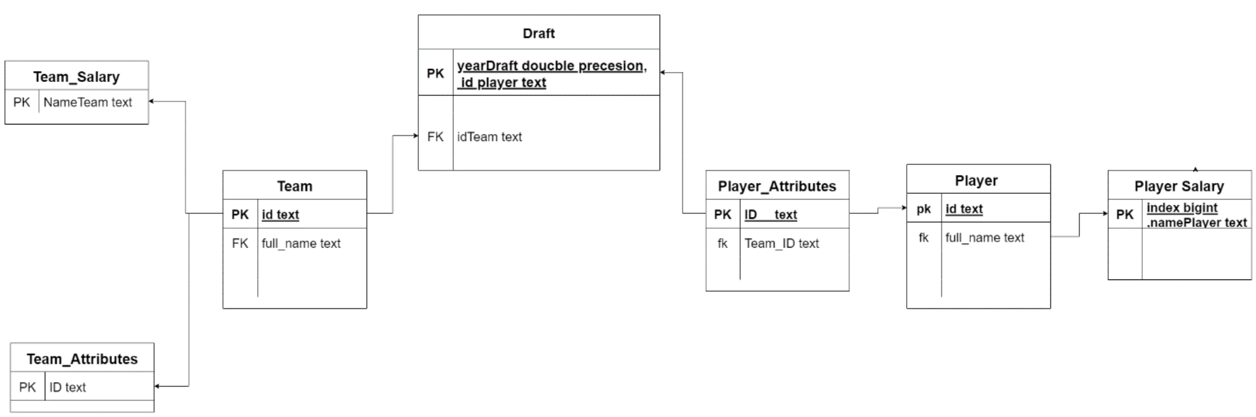
3.0 Database

3.1 ERD of RDBMS (basket_database)

Our RDBMS in this project, also known as basket_database, contains 16 tables. To present basket_database in a relational model and relationship between entities, we will use the Entity–Relationship Model (ERD). Because few tables have more than 50 columns or attributes, it is not possible to present as a figure on this page. The ERD diagrams will be available through a Google Drive link in Appendix 1. Please take note that the ERD was built using PostgreSQL, and that PostgreSQL also auto aligned the ERD so that it appears in the correct order.

3.2 ERD of RDBMS (basket_database)

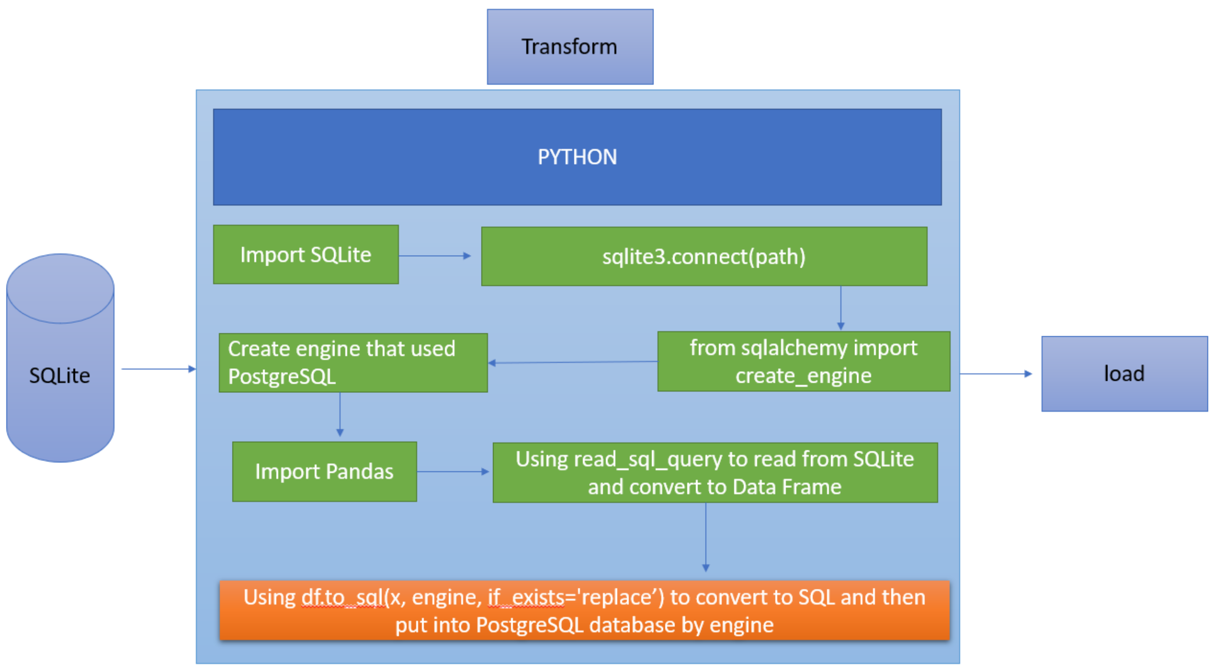
Our data warehouse in this project, also known as basket_dw, contains 7 tables after transformed. We also used draw.io and PostgreSQL build-in tool to present data warehouse schema. We will use draw.io to illustrate how data warehouse schema will look like as figure below. But only the Primary key (PK), Foreign key (FK) and table name will be display since there are some tables have more than 30 columns. The overall pictures will be available through a Google drive link in appendix 2. As we can see in figure below, it is a galaxy schema. This is because Team and Player_Attributes are fact table, while for Team_Salary, Team_Attributes, Draft, Player, and Player_Salary table are dimension table.



4.0 ETL Pipeline

My sincere gratitude goes out to Leong Teng Man for advising me in the use of Python for the ETL process.

4.1 ETL (from data source to RDBMS)

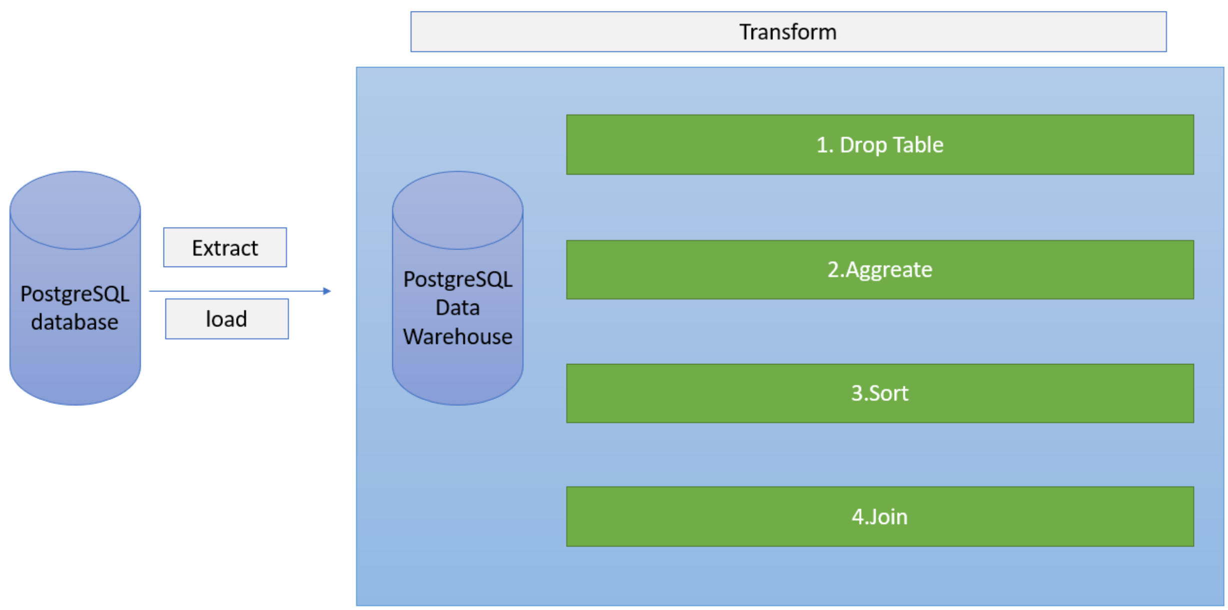


SQLite is a lightweight SQL embedded database engine, as stated in the 2.0 Architecture. Due to the fact that SQLite and PostgreSQL implement distinct data types, which cannot be directly exchanged with each other. We will use python to perform ETL before the data goes into database. The code for the ETL process will be provided in appendix 3. The step will be as follows:

1. Download and import SQLite, SQLAlchemy and pandas library.
2. Use connect function that provided by SQLite python library to connect the SQLite database.
3. Run a query in python that displays all accessible tables to see what kinds of tables SQLite has and how many of each kind there are. (query= "SELECT name FROM sqlite_master WHERE type='table';")
4. Create an engine that based on PostgreSQL for the next step.

5. Use read_sql_query() function provided by pandas to read data from the SQLite database, and then convert into pandas data frame
6. In the final step, the data frame is converted to SQL using the dataframe.to sql() function, and then the engine created in step 4 is used to insert data into the database by running SQL operations.

4.2 ELT (from RDBMS to Data Warehosue)



Since we use the same platform (PostgreSQL) for both RDBMS and data warehousing. So, we don't have to transform the data before loading it into a data warehouse. This method saves a lot of time and is easy to change. We can transform data whenever we need to take a transformation step. In fact, ETL is needed because business data in modern computing is stored in many different places and formats that don't work together. Anyways, we have no SQL database for it.

When it was necessary, various transformation procedures, including drop table, sort table, join table, aggregate table, and Lookup, were applied. Example of code for ELT will be available in appendix 4.The following will serve as a description of the method of transformation:

1. Drop table

   - to remove a table and all its rows
   - we succeed dropped 9 tables for OLAP in data warehouse, basket_dw

2. Aggregate table

   - Creating aggregate tables reduces data warehouse access. More than one aggregate table can improve schema performance. (MIN, MAX,SUM,COUNT,AVG and so on)

3. Joining table

   - Joining multiple attributes into one. (Inner join, outer join, full join and so on)

4. Sorting

   - sorting data or row data or tuple based on the basis of some attribute. (Order by)

5.0 Results and Data Analysis

Tableau excels in working with relational databases, and while it can connect to OLAP cubes, doing so duplicates Tableau's aggregation and hierarchy features. Tableau is also suited for a variety of non-SQL and big data solutions, such as the direct Google Big Query connector.

5.1 The 2D and 3D representation data

i) Performance and Salary of Players

Example of 2D factory data with only 3 teams included

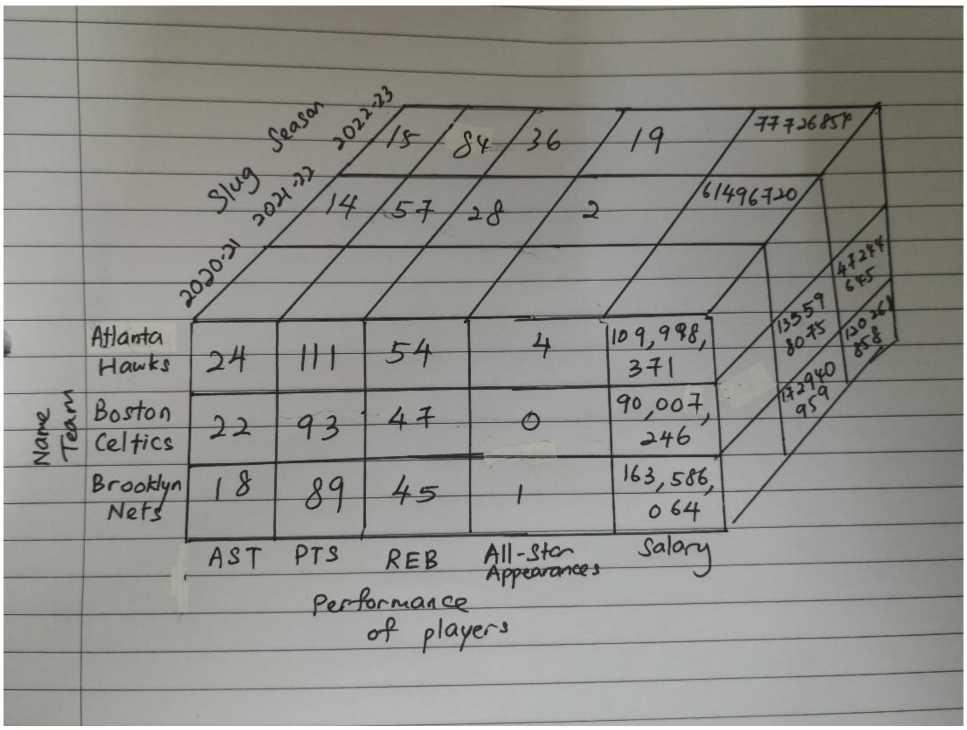
(Only show few of the data since there is impossible to show whole data)

Slug Season = 2020-21					
	Performance of Players				
Name Team	AST	PTS	REB	All Star Appearances	Salary
Atlanta Hawks	24	111	54	4	109998371
Boston Celtics	22	93	47	0	90007246
Brooklyn Nets	18	89	45	1	163586064
Slug Season = 2021-22					
	Performance of Players				
Name Team	AST	PTS	REB	All Star Appearances	Salary
Atlanta Hawks	14	57	28	2	61496720
Boston Celtics	17	84	35	10	135598075
Brooklyn Nets	17	95	40	9	172940959
Slug Season = 2022-23					
	Performance of Players				
Name Team	AST	PTS	REB	All Star Appearances	Salary
Atlanta Hawks	15	84	36	19	77726854
Boston Celtics	6	25	12	0	47244645
Brooklyn Nets	24	83	13	9	120268858

Example of 3D data representation as 2D

Name Team	Slug Season = 2020-21					Slug Season = 2021-22					Slug Season = 2022-23				
	Performance of Players					Performance of Players					Performance of Players				
	AST	PTS	REB	All Star Appearances	Salary	AST	PTS	REB	All Star Appearances	Salary	AST	PTS	REB	All Star Appearances	Salary
Atlanta Hawks	24	111	54	4	109998371	14	57	28	2	61496720	15	84	36	19	77726854
Boston Celtics	22	93	47	0	90007246	17	84	35	10	135598075	6	25	12	0	47244645
Brooklyn Nets	18	89	45	1	163586064	17	95	40	9	172940959	24	83	13	9	120268858

3D Olap Cube



ii) Experience of Players

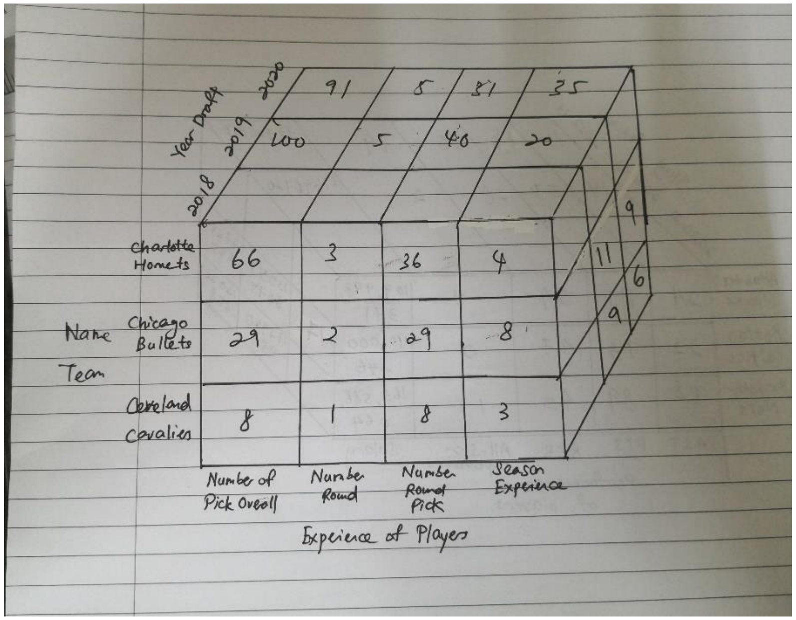
Example of 2D factory data with only 3 teams included

Year Draft = 2018				
Name Team	Experience of Players			
	Number Pick Overall	Number Round	Number Round Pick	Season Experience
Charlotte Hornets	66	3	36	4
Chicago Bullets	29	2	29	8
Cleveland Cavaliers	8	1	8	3
Year Draft = 2019				
Name Team	Experience of Players			
	Number Pick Overall	Number Round	Number Round Pick	Season Experience
Charlotte Hornets	100	5	40	20
Chicago Bullets	45	3	15	11
Cleveland Cavaliers	31	2	31	9
Year Draft = 2020				
Name Team	Experience of Players			
	Number Pick Overall	Number Round	Number Round Pick	Season Experience
Charlotte Hornets	91	5	31	35
Chicago Bullets	48	3	18	9
Cleveland Cavaliers	5	1	5	6

Example of 3D data representation as 2D

	Year Draft = 2018				Year Draft = 2019				Year Draft = 2020			
	Experience of Players				Experience of Players				Experience of Players			
Name Team	Number Pick Overall	Number Round	Number Round Pick	Season Experience	Number Pick Overall	Number Round	Number Round Pick	Season Experience	Number Pick Overall	Number Round	Number Round Pick	Season Experience
Charlotte Hornets	66	3	36	4	100	5	40	20	91	5	31	35
Chicago Bullets	29	2	29	8	45	3	15	11	48	3	18	9
Cleveland Cavaliers	8	1	8	3	31	2	31	9	5	1	5	6

3D Olap Cube



5.2 Olap Operations

i) Performance and Salary of Players

First, show the original table of olap operations

(All pictures below only show a few of data but the whole data)

(Whole data can be seen in the tableau that is provided)

Slug Sea..	Name Team	Full Name	AST	PTS	REB	All Star Appearances	Salary
2020-21	Atlanta Hawks	Alaa Abdelnaby	0	6	3	0	18,000,000
		Alexis Ajinca	1	5	4	0	5,813,640
		Bam Adebayo	5	19	10		4,137,302
		Charles Acton	1	3	2	0	16,000,000
		Danny Ainge	4	12	3	1	8,000,000
		Don Adams	2	9	6	0	19,500,000
		Furkan Aldemir	1	2	4	0	6,571,800
		Gary Alcorn	0	3	4	0	12,178,571
		Jordan Adams	1	3	1	0	7,422,000
		Mark Aguirre	3	20	5	3	4,767,000
		Mark Alarie	1	8	3	0	1,620,564
		Rick Adelman	4	8	2	0	2,761,920
		Tariq Abdul-Wahad	1	8	3	0	1,701,593
		Tom Abernethy	1	6	3	0	1,517,981
	Boston Celtics	Al-Farouq Aminu	2	5	5		2,250,000
		Chuck Aleksinas	1	5	4	0	3,458,400
		Derek Anderson	3	12	3	0	2,029,920
		Grayson Allen	2	10	3		17,450,000
		Greg Anderson	1	7	6	0	3,631,200
		Jarrett Allen	2	13	10		2,498,760
		Joe Alexander	1	4	2	0	1,517,981
		Lance Allred	0	1	0	0	9,897,120
		Lou Amundson	0	4	4	0	13,446,428
		Lucius Allen	5	13	3	0	1,039,080
		Michael Anderson	4	6	3	0	9,258,000
		Nickeil Alexander-W..	2	9	3		92,857
		Randy Allen	0	4	2	0	23,437,500
	Brooklyn Nets	B.J. Armstrong	3	10	2	1	263,995
		Bob Armstrong	0	2	2	0	217,845
		Brandon Armstrong	0	2	1	0	11,454,048
		Hilton Armstrong	0	3	3	0	1,824,003
		Kostas Antetokoun..	0	1	2		1,620,564
		Michael Ansley	0	7	4	0	41,254,920

1. Roll-up operation

Rolling up from player to team.

- By using roll-up operations, we can see which slug season has which team that has highest Assist (AST), Points Scored (PTS), Rebounds (REB), All Star Appearances and most rational salary.

Slug Sea..	Name Team	AST	PTS	REB	All Star Appearances	Salary
2020-21	Atlanta Hawks	24	111	54	4	109,992,371
	Boston Celtics	22	93	47	0	90,007,246
	Brooklyn Nets	18	89	45	1	163,586,064
	Charlotte Hornets	26	86	39	0	107,320,579
	Chicago Bulls	27	117	54	2	112,674,509
	Cleveland Cavaliers	34	164	63	9	125,263,154
	Dallas Mavericks	16	85	50	6	104,103,762
	Denver Nuggets	30	124	49	15	128,279,712
	Detroit Pistons	21	104	56	1	104,972,497
	Golden State Warrio..	20	95	49	9	168,750,607
	Houston Rockets	37	155	54	8	123,102,045
	Indiana Pacers	17	85	26	1	116,412,709
	Los Angeles Clippers	16	68	34	0	136,674,513
	Los Angeles Lakers	20	95	44	1	121,227,721
	Memphis Grizzlies	25	118	60	1	105,420,763
	Miami Heat	27	100	46	2	131,879,929
	Milwaukee Bucks	19	64	27	4	89,513,837
	Minnesota Timberw..	23	86	29	3	123,560,783
	New Orleans Pelicans	23	79	44	5	101,048,939
	New York Knicks	18	83	35	0	67,525,105
	Oklahoma City Thun..	30	125	38	6	46,105,073
	Orlando Magic	22	107	45	7	111,586,242
	Philadelphia 76ers	17	54	29	0	91,483,027
	Phoenix Suns	26	104	48	3	122,756,196
	Portland Trail Blazers	23	102	40	10	124,497,534
	Sacramento Kings	25	84	41	15	102,271,537
	San Antonio Spurs	19	111	53	4	121,181,500
	Toronto Raptors	14	73	36	0	99,897,736
	Utah Jazz	11	57	30	0	133,955,577
	Washington Wizards	23	111	49	7	129,538,264
2021-22	Atlanta Hawks	14	57	28	2	61,496,720
	Boston Celtics	17	84	35	10	135,598,075
	Brooklyn Nets	17	95	40	9	172,940,959

2. Drill down operation

Drill down from team to player.

- By using drill-down operation, we can see which slug season has which player has highest AST, PTS, REB, All Star Appearances and most rational salary

Slug Sea..	Full Name	AST	PTS	REB	All Star Appearances	Salary
2021-22	Tony Bradley	1	5	5		666,667
	Tony Campbell	2	12	3	0	4,054,695
	Tony Dawson	0	3	1	0	2,303,040
	Torrey Craig	1	3	2		6,431,667
	Travis Diener	2	5	1	0	2,602,920
	Troy Brown Jr.	1	4	3		7,518,518
	Tyler Cavanaugh	1	4	3	0	2,711,280
	Tyler Ennis	2	4	1	0	1,517,981
	Vin Baker	2	15	7	4	10,183,800
	Vitor Faverani	0	4	4	0	10,851,246
	Walter Dukes	1	10	11	2	11,615,328
	Wayne Englestad	1	3	2	0	2,959,080
	Wesley Cox	0	5	3	0	1,669,178
	William Bell	2	5	3	0	11,615,328
	Willie Anderson	4	12	4	0	1,729,217
	Wilt Chamberlain	4	30	23	13	4,004,280
	Winston Bennett	1	5	3	0	3,150,000
	Winston Crite	1	3	2	0	5,105,160
	Zaid Abdul-Aziz	1	9	8	0	18,000,000
2022-23	A.J. English	2	10	2	0	4,556,983
	Aaron Brooks	3	10	2	0	122,741
	Al Bianchi	2	8	3	0	4,000,000
	Al Carlson	1	3	3	0	2,300,000
	Al Ferrari	3	7	2	0	13,340,000
	Al Fleming	0	2	2	0	2,174,880
	Alex Blackwell	0	1	1	0	2,228,276
	Alexander Ellis	1	5	5	0	4,437,000
	Alvan Adams	4	14	7	1	18,206,896
	Anthony Avent	1	6	5	0	8,623,920
	Anthony Brown	1	4	3	0	18,796,296
	Anthony Davis	3	23	8		1,900,000
	Antonio Burks	1	2	1	0	2,165,298
	Antonio Daniels	3	8	2	0	999,200

3. Slicing

Slicing the data for slug season 2020-21.

- By using slicing operation, we are able to see the performance and salary of team and player by specific slug season

Slug Sea..	Name Te..	Full Name	AST	PTS	REB	All Star Appearances	Salary
2020-21	Atlanta Hawks	Alaa Abdelnaby	0	6	3	0	18,000,000
		Alexis Ajinca	1	5	4	0	5,813,640
		Bam Adebayo	5	19	10		4,137,302
		Charles Acton	1	3	2	0	16,000,000
		Danny Ainge	4	12	3	1	8,000,000
		Don Adams	2	9	6	0	19,500,000
		Furkan Aldemir	1	2	4	0	6,571,800
		Gary Alcorn	0	3	4	0	12,178,571
		Jordan Adams	1	3	1	0	7,422,000
		Mark Aguirre	3	20	5	3	4,767,000
		Mark Alarie	1	8	3	0	1,620,564
		Rick Adelman	4	8	2	0	2,761,920
		Tariq Abdul-Wahad	1	8	3	0	1,701,593
		Tom Abernethy	1	6	3	0	1,517,981
	Boston Celtics	Al-Farouq Aminu	2	5	5		2,250,000
		Chuck Aleksinas	1	5	4	0	3,458,400
		Derek Anderson	3	12	3	0	2,029,920
		Grayson Allen	2	10	3		17,450,000
		Greg Anderson	1	7	6	0	3,631,200
		Jarrett Allen	2	13	10		2,498,760
		Joe Alexander	1	4	2	0	1,517,981
		Lance Allred	0	1	0	0	9,897,120
		Lou Amundson	0	4	4	0	13,446,428
		Lucius Allen	5	13	3	0	1,039,080
		Michael Anderson	4	6	3	0	9,258,000
		Nickeil Alexander-W..	2	9	3		92,857
		Randy Allen	0	4	2	0	23,437,500
	Brooklyn Nets	B.J. Armstrong	3	10	2	1	263,995
		Bob Armstrong	0	2	2	0	217,845
		Brandon Armstrong	0	2	1	0	11,454,048
		Hilton Armstrong	0	3	3	0	1,824,003
		Kostas Antetokoun..	0	1	2		1,620,564
		Michael Ansley	0	7	4	0	41,254,920

Slicing the data from slug season only with name team.

Slug Sea..	Name Team	AST	PTS	REB	All Star Appearances	Salary
2020-21	Atlanta Hawks	24	111	54	4	109,992,371
	Boston Celtics	22	93	47	0	90,007,246
	Brooklyn Nets	18	89	45	1	163,586,064
	Charlotte Hornets	26	86	39	0	107,320,579
	Chicago Bulls	27	117	54	2	112,674,509
	Cleveland Cavaliers	34	164	63	9	125,263,154
	Dallas Mavericks	16	85	50	6	104,103,762
	Denver Nuggets	30	124	49	15	128,279,712
	Detroit Pistons	21	104	56	1	104,972,497
	Golden State Warrio..	20	95	49	9	168,750,607
	Houston Rockets	37	155	54	8	123,102,045
	Indiana Pacers	17	85	26	1	116,412,709
	Los Angeles Clippers	16	68	34	0	136,674,513
	Los Angeles Lakers	20	95	44	1	121,227,721
	Memphis Grizzlies	25	118	60	1	105,420,763
	Miami Heat	27	100	46	2	131,879,929
	Milwaukee Bucks	19	64	27	4	89,513,837
	Minnesota Timberw..	23	86	29	3	123,560,783
	New Orleans Pelicans	23	79	44	5	101,048,939
	New York Knicks	18	83	35	0	67,525,105
	Oklahoma City Thun..	30	125	38	6	46,105,073
	Orlando Magic	22	107	45	7	111,586,242
	Philadelphia 76ers	17	54	29	0	91,483,027
	Phoenix Suns	26	104	48	3	122,756,196
	Portland Trail Blazers	23	102	40	10	124,497,534
	Sacramento Kings	25	84	41	15	102,271,537
	San Antonio Spurs	19	111	53	4	121,181,500
	Toronto Raptors	14	73	36	0	99,897,736
	Utah Jazz	11	57	30	0	133,955,577
	Washington Wizards	23	111	49	7	129,538,264

4.Dicing

Dicing the data of slug season 2021-22 and name team is Charlotte Hornets with players.

- By using dicing operation, we are able to know the performance and salary of the specific slug season and name team

Slug Sea..	Name Te..	Full Name	AST	PTS	REB	All Star Appearances	Salary
2021-22	Charlotte Hornets	Alvin Attles	4	9	4	0	29,925,000
		Bob Arnzen	0	6	3	0	2,812,500
		Cameron Bairstow	0	1	1	0	1,517,981
		Darrell Arthur	1	7	4	0	1,782,621
		Deandre Ayton	2	15	11		5,421,493
		James Augustine	0	2	1	0	1,782,621
		Jim Baechtold	2	10	3	0	9,043,478
		Ken Austin	0	2	0	0	8,231,760
		Luke Babbitt	1	5	2	0	1,517,981
		Marvin Bagley III	1	14	7		4,215,120
		Omer Asik	1	5	7	0	1,782,621
		Richard Atha	1	3	2	0	4,736,102
		Thurl Bailey	1	13	5	0	17,905,263

(Dicing the data of slug season 2021-22 and name team is Charlotte Hornets)

Slug Sea..	Name Team	AST	PTS	REB	All Star Appearances	Salary
2021-22	Charlotte Hornets	14	90	50	0	90,674,541

5. Pivot

Pivot the diced data of slug season 2021-22 and name team is Charlotte Hornets with players.

- Pivot the data to see in alternative way

	Slug Season / Name Team / Full Name 2021-22 Charlotte Hornets												
	Alvin Attles	Bob Arnzen	Cameron B..	Darrell Art..	Deandre Ay..	James Aug..	Jim Baecht..	Ken Austin	Luke Babbitt	Marvin Bag..	Omer Asik	Richard Atha	Thurl Bailey
AST	4	0	0	1	2	0	2	0	1	1	1	1	1
PTS	9	6	1	7	15	2	10	2	5	14	5	3	13
REB	4	3	1	4	11	1	3	0	2	7	7	2	5
All Star A..	0	0	0	0		0	0	0	0		0	0	0
Salary	29,925,000	2,812,500	1,517,981	1,782,621	5,421,493	1,782,621	9,043,478	8,231,760	1,517,981	4,215,120	1,782,621	4,736,102	17,905,263

ii) Experience of Players

First, show the original table of olap operations

Year Draft	Name Team	Full Name	Number Pick Overall	Number Round	Number Round Pick	Season Exp
1974	Atlanta Hawks	Charlie Yelverton	7.0	1.0	7.0	0.0
		Danny Young	10.0	1.0	10.0	9.0
		Gary Zeller	25.0	2.0	7.0	1.0
	Boston Celtics	Sam Young	17.0	1.0	17.0	4.0
		Stephen Zimmerm..	35.0	2.0	17.0	1.0
	Buffalo Braves	Rich Yonakor	9.0	1.0	9.0	0.0
	Chicago Bulls	Michael Young	14.0	1.0	14.0	2.0
		Perry Young	16.0	1.0	16.0	0.0
	Cleveland Cavaliers	Bill Zopf	39.0	3.0	3.0	0.0
		Jim Zoet	38.0	3.0	2.0	0.0
		Yi Jianlian	8.0	1.0	8.0	5.0
	Golden State Warrio..	Tony Zeno	29.0	2.0	11.0	0.0
	Houston Rockets	Barry Yates	5.0	1.0	5.0	0.0
		Cody Zeller	23.0	2.0	5.0	7.0
	Kansas City-Omaha Kings	Dave Zeller	24.0	2.0	6.0	0.0
		Wayne Yates	6.0	1.0	6.0	0.0
	Los Angeles Lakers	Joe Young	12.0	1.0	12.0	2.0
	Milwaukee Bucks	Thaddeus Young	18.0	1.0	18.0	13.0
	New Orleans Jazz	Tyler Zeller	28.0	2.0	10.0	8.0
	Philadelphia 76ers	Tim Young	19.0	2.0	1.0	0.0
		Yao Ming	2.0	1.0	2.0	8.0
	Phoenix Suns	George Yardley	4.0	1.0	4.0	6.0
		Ivica Zubac	40.0	3.0	4.0	4.0
	Portland Trail Blazers	Paul Zipser	36.0	2.0	18.0	1.0
		Trae Young	20.0	2.0	2.0	2.0
	Seattle SuperSonics	Vincent Yarbrough	3.0	1.0	3.0	1.0
	Washington Bullets	Phil Zevenbergen	30.0	2.0	12.0	0.0
		Robert Zawoluk	22.0	2.0	4.0	2.0
1975	Atlanta Hawks	Andrew Wiggins	1.0	1.0	1.0	6.0
		Ken Wilburn	3.0	1.0	3.0	1.0
		Qyntel Woods	146.0	9.0	2.0	3.0
		Rick Wilson	111.0	7.0	3.0	1.0
		Robert Williams III	93.0	6.0	3.0	2.0

1. Roll-up

Rolling up from player to team.

- By using roll-up operations, we can see which team has the most number pick overall, number round, number round pick, and season experience.

Year Draft	Name Team	Number Pick Overall	Number Round	Number Round Pick	Season Exp
1974	Atlanta Hawks	42	4	24	10
	Boston Celtics	52	3	34	5
	Buffalo Braves	9	1	9	0
	Chicago Bulls	30	2	30	2
	Cleveland Cavaliers	85	7	13	5
	Golden State Warrio..	29	2	11	0
	Houston Rockets	28	3	10	7
	Kansas City-Omaha ..	30	3	12	0
	Los Angeles Lakers	12	1	12	2
	Milwaukee Bucks	18	1	18	13
	New Orleans Jazz	28	2	10	8
	Philadelphia 76ers	21	3	3	8
	Phoenix Suns	44	4	8	10
	Portland Trail Blazers	56	4	20	3
	Seattle SuperSonics	3	1	3	1
	Washington Bullets	52	4	16	2
1975	Atlanta Hawks	483	32	15	13
	Boston Celtics	941	53	151	40
	Buffalo Braves	772	44	90	17
	Chicago Bulls	476	28	80	39
	Cleveland Cavaliers	878	54	106	61
	Detroit Pistons	655	40	63	24
	Golden State Warrio..	610	36	106	46
	Houston Rockets	616	37	76	59
	Kansas City Kings	569	34	101	42
	Los Angeles Lakers	388	26	10	15
	Milwaukee Bucks	785	49	49	29
	New Orleans Jazz	679	45	15	22
	New York Knicks	930	57	104	29
	Philadelphia 76ers	700	45	36	39
	Phoenix Suns	815	51	79	35
	Portland Trail Blazers	779	48	43	37
	Seattle SuperSonics	664	40	72	28

2. Drill-down operation

Drill down from team to player.

- By using drill-down operations, we can see which player has the most number pick overall, number round, number round pick, and season experience

Year Draft	Full Name	Number Pick Overall	Number Round	Number Round Pick	Season Exp
1974	Barry Yates	5.0	1.0	5.0	0.0
	Bill Zopf	39.0	3.0	3.0	0.0
	Charlie Yelverton	7.0	1.0	7.0	0.0
	Cody Zeller	23.0	2.0	5.0	7.0
	Danny Young	10.0	1.0	10.0	9.0
	Dave Zeller	24.0	2.0	6.0	0.0
	Gary Zeller	25.0	2.0	7.0	1.0
	George Yardley	4.0	1.0	4.0	6.0
	Ivica Zubac	40.0	3.0	4.0	4.0
	Jim Zoet	38.0	3.0	2.0	0.0
	Joe Young	12.0	1.0	12.0	2.0
	Michael Young	14.0	1.0	14.0	2.0
	Paul Zipser	36.0	2.0	18.0	1.0
	Perry Young	16.0	1.0	16.0	0.0
	Phil Zevenbergen	30.0	2.0	12.0	0.0
	Rich Yonakor	9.0	1.0	9.0	0.0
	Robert Zawoluk	22.0	2.0	4.0	2.0
	Sam Young	17.0	1.0	17.0	4.0
	Stephen Zimmerm..	35.0	2.0	17.0	1.0
	Thaddeus Young	18.0	1.0	18.0	13.0
	Tim Young	19.0	2.0	1.0	0.0
	Tony Zeno	29.0	2.0	11.0	0.0
	Trae Young	20.0	2.0	2.0	2.0
	Tyler Zeller	28.0	2.0	10.0	8.0
	Vincent Yarbrough	3.0	1.0	3.0	1.0
	Wayne Yates	6.0	1.0	6.0	0.0
	Yao Ming	2.0	1.0	2.0	8.0
	Yi Jianlian	8.0	1.0	8.0	5.0
1975	A.J. Wynder	174.0	10.0	14.0	0.0
	Al Wood	139.0	8.0	13.0	5.0
	Alan Williams	24.0	2.0	6.0	4.0
	Andrew Wiggins	1.0	1.0	1.0	6.0
	Arthur Williams	26.0	2.0	8.0	6.0

3. Slicing

Slicing the data from year draft 2020.

- By using slicing operation, we are able to get experience of player in specific year draft

Year D...	Name Team	Full Name	Number Pick Overall	Number Round	Number Round Pick	Season Exp
2020	Atlanta Hawks	Shareef Abdur-Rah..	6.00	1.00	6.00	12.00
		Victor Alexander	50.00	2.00	20.00	4.00
	Boston Celtics	Charles Acton	14.00	1.00	14.00	0.00
		Gary Alexander	47.00	2.00	17.00	0.00
		Mark Aguirre	30.00	1.00	30.00	12.00
		Rick Adelman	26.00	1.00	26.00	6.00
	Brooklyn Nets	Grayson Allen	55.00	2.00	25.00	2.00
	Charlotte Hornets	Danny Ainge	32.00	2.00	2.00	13.00
		Jarrett Allen	56.00	2.00	26.00	3.00
		Kareem Abdul-Jab..	3.00	1.00	3.00	19.00
	Chicago Bulls	Cliff Alexander	44.00	2.00	14.00	1.00
		Mahmoud Abdul-R..	4.00	1.00	4.00	8.00
	Cleveland Cavaliers	Tariq Abdul-Wahad	5.00	1.00	5.00	6.00
	Dallas Mavericks	Blake Ahearn	31.00	2.00	1.00	3.00
	Denver Nuggets	Steven Adams	22.00	1.00	22.00	7.00
	Detroit Pistons	Tom Abernethy	7.00	1.00	7.00	4.00
	Golden State Warriors	Joe Alexander	48.00	2.00	18.00	2.00
		Nickeil Alexander-..	51.00	2.00	21.00	1.00
		Zaid Abdul-Aziz	2.00	1.00	2.00	9.00
	Indiana Pacers	Bob Allen	54.00	2.00	24.00	0.00
	LA Clippers	Jerome Allen	57.00	2.00	27.00	1.00
	Los Angeles Lakers	Arron Afflalo	28.00	1.00	28.00	11.00
	Memphis Grizzlies	Furkan Aldemir	40.00	2.00	10.00	1.00
	Miami Heat	Jordan Adams	20.00	1.00	20.00	2.00
	Milwaukee Bucks	Bam Adebayo	24.00	1.00	24.00	3.00
		Cory Alexander	45.00	2.00	15.00	7.00
	Minnesota Timberwolves	Alaa Abdelnaby	1.00	1.00	1.00	4.00
		Alexis Ajinca	33.00	2.00	3.00	7.00
		Don Adams	17.00	1.00	17.00	6.00
	New Orleans Pelicans	Gary Alcorn	39.00	2.00	9.00	1.00
		LaMarcus Aldridge	42.00	2.00	12.00	14.00
		Lucius Allen	60.00	2.00	30.00	9.00
		Mark Acres	13.00	1.00	13.00	5.00

4. Dicing

Dicing the data from year draft 2020 and Name team is Atlanta Hawks.

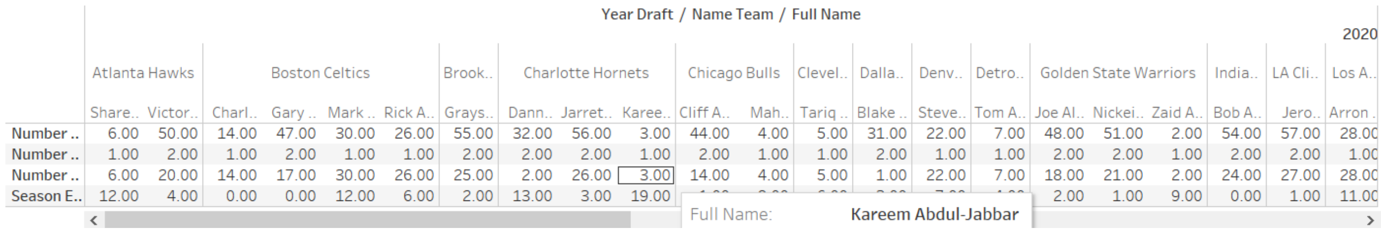
- By using dicing operation, we are able to know the experience of players from specific year draft and specific team

Year Draft	Name Te..	Full Name	Number Pick Overall	Number Round	Number Round Pick	Season Exp
2020	Atlanta Hawks	Shareef Abdur-Rah..	6.00	1.00	6.00	12.00
		Victor Alexander	50.00	2.00	20.00	4.00

5. Pivot

Pivot the slicing data from year draft 2020.

- Pivot the data to see in alternative way



Year Draft / Name Team / Full Name

2020

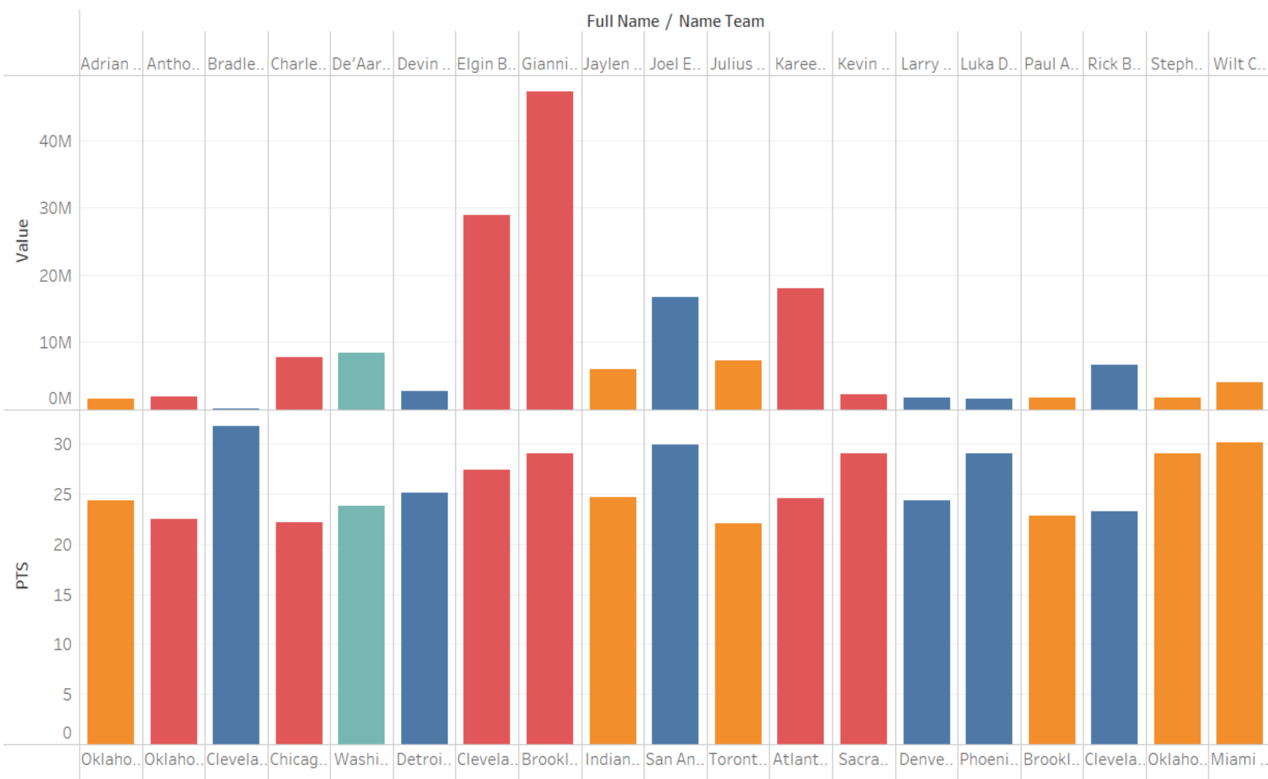
	Atlanta Hawks		Boston Celtics				Brook..	Charlotte Hornets			Chicago Bulls	
	Share..	Victor..	Charl..	Gary ..	Mark ..	Rick A..	Grays..	Dann..	Jarret..	Karee..	Cliff A..	Mah..
Number ..	6.00	50.00	14.00	47.00	30.00	26.00	55.00	32.00	56.00	3.00	44.00	4.00
Number ..	1.00	2.00	1.00	2.00	1.00	1.00	2.00	2.00	2.00	1.00	2.00	1.00
Number ..	6.00	20.00	14.00	17.00	30.00	26.00	25.00	2.00	26.00	3.00	14.00	4.00
Season E..	12.00	4.00	0.00	0.00	12.00	6.00	2.00	13.00	3.00	19.00	[illegible]	[illegible]

	Clevel..	Dalla..	Denv..	Detro..	Golden State Warriors			India..	LA Cli..	Los A..
	Tariq ..	Blake ..	Steve..	Tom A..	Joe Al..	Nickei..	Zaid A..	Bob A..	Jero..	Arron ..
Number ..	5.00	31.00	22.00	7.00	48.00	51.00	2.00	54.00	57.00	28.00
Number ..	1.00	2.00	1.00	1.00	2.00	2.00	1.00	2.00	2.00	1.00
Number ..	5.00	1.00	22.00	7.00	18.00	21.00	2.00	24.00	27.00	28.00
Season E..	[illegible]	[illegible]	[illegible]	[illegible]	2.00	1.00	9.00	0.00	1.00	11.00

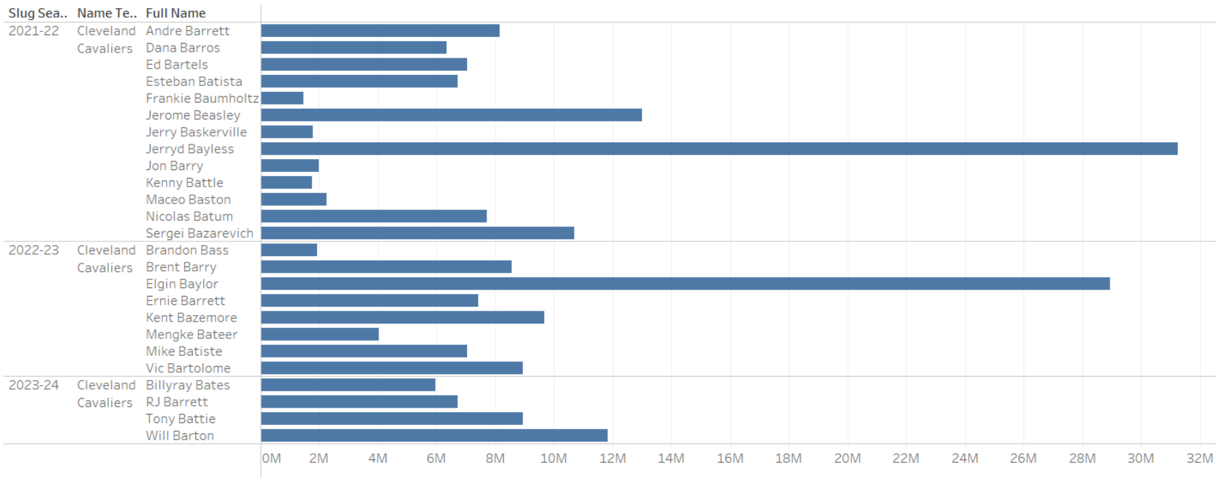
Full Name: Kareem Abdul-Jabbar

5.3 Data Visualization

In the first OLAP operation, we found that there is too much data. Therefore, in order to find out who is the most potential as well as less budget can be covered. We must filter it first. Therefore, the graph only shows the player that achieved 22.00 PTS (points scored per game) so that is easy for us to see who is the most potential.

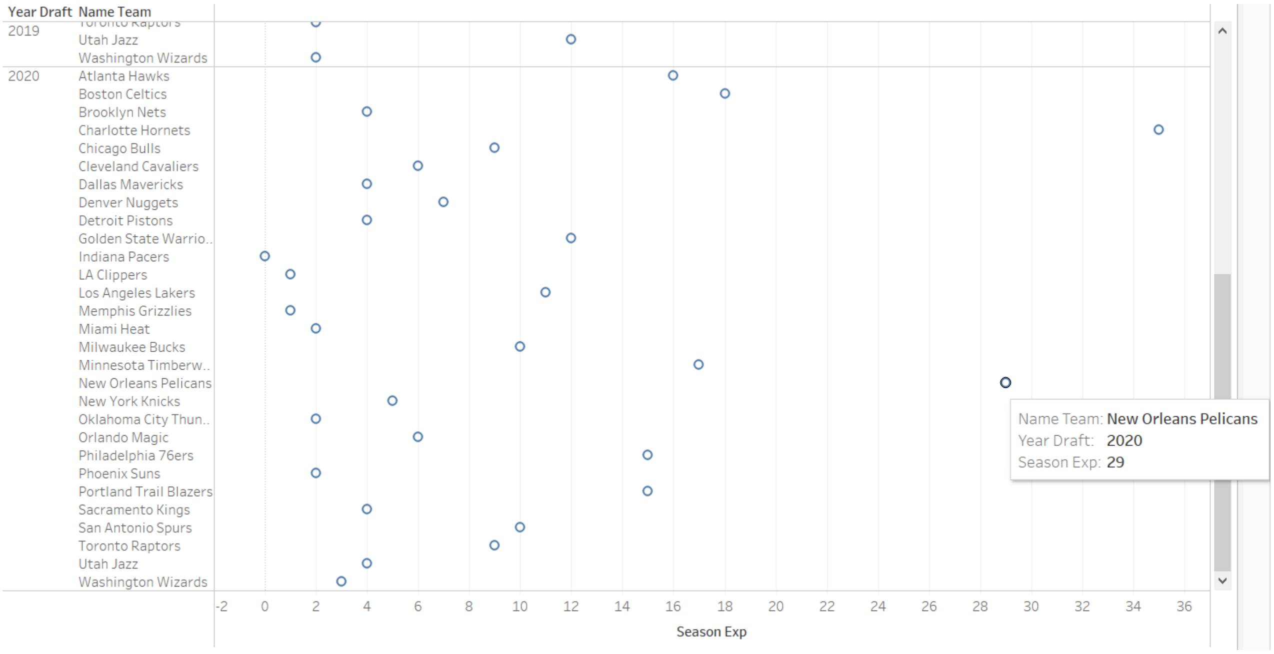


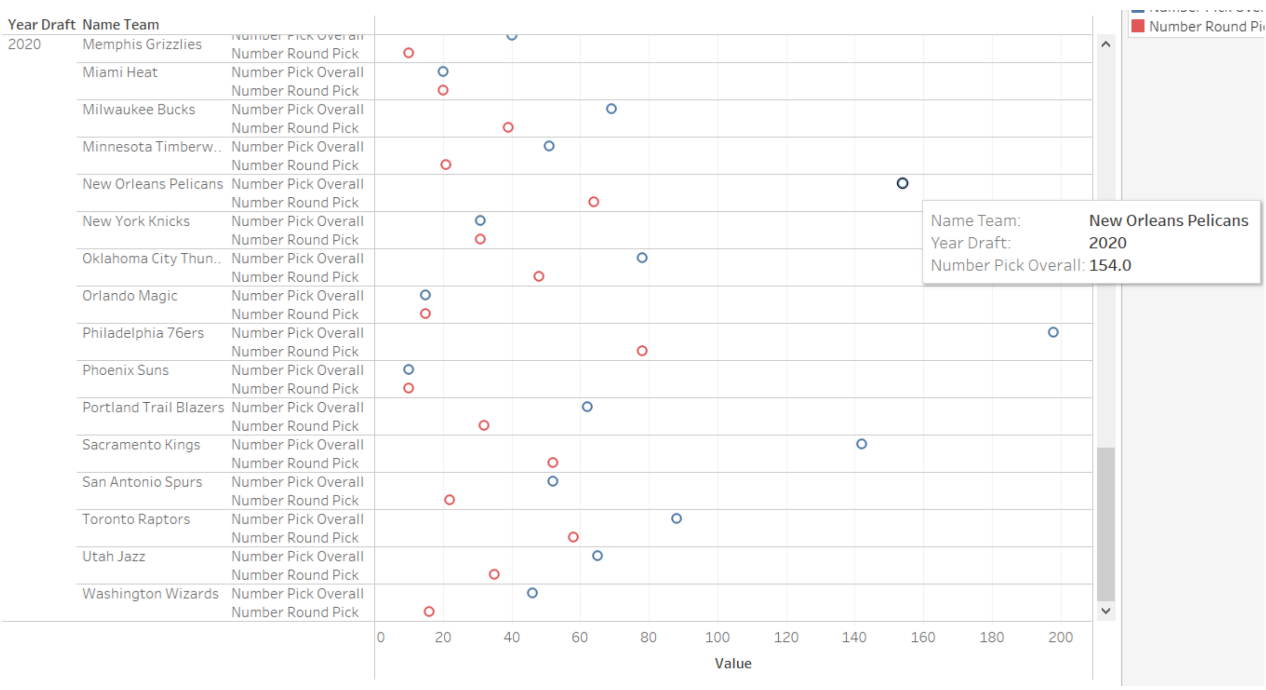
The value means salary required for the player while PTS is point scored. We can see that one player, Bradley Beal, required the lowest salary but achieved the highest score in PTS. It means that he has got the highest potential among the players with a low budget.



After the slug season 2020, the player is not shown in the graph. Perhaps the team, Cleveland Cavaliers did not pay attention to him. Therefore, with less budget and if we require high performance player, Bradley Beal is the best choice regarding the OLAP operations and data.

According to second OLAP operations, considering our budget is enough, we tried to find out which team has the most potential to be bought.





Based on these two graphs, New Orleans Pelicans has the highest number pick overall and season experience. Therefore, it is the team that contains the highest potential. It is because pick overall means that the team is facing a losing period and it has the priority to choose the player from draft with highest skills and scored. Not only that, it also has many season experience. Experience in the field is much important than the skills. Therefore, it is quite worth to be selected.

6.0 Conclusion

In the end, we were able to come to the conclusion that Bradley Beal, who plays for the Cleveland Cavaliers, is the player who is both the most cost effective and the most underrated that we discovered using OLAP analysis in 2021. Since the New Orleans Pelicans had the highest pick among all teams, this indicates that they are the most likely team to be likely to succeed and potential.

One of the challenges is we must use a database as part of analysis. Because data in modern computing is stored in a wide variety of locations and formats, which, with a few exceptions, are incompatible with one another, we need to construct a complicated ETL pipeline.

In certain circumstances, a non-SQL database is a much better option than an RDBMS due to the complex ETL process and the requirement of a pre-defined format. non-SQL database support for machines Learning, such as that provided by MangoDB, is significantly more useful than OLAP analysis.

PostgreSQL is not a suitable option for use as a data warehouse over the long term. The price of other warehouses, such as Snowflake and Red Shift, can range anywhere from two hundred thousand dollars to two million dollars. According to Roi Avinoam (2020), the answer to this question depends on the intricacy and specificity of your data, the size of your organization, whether or not you build the data warehouse from scratch as opposed to purchasing a Data Warehouse as a Service (DWaaS), and various other factors such as pricing structures.

Due to the fact that the formats, data structures, and logical schemas of various databases are all distinct from one another. The process of migrating data from RDBMS to data warehouse is limited by the lack of available resources and free solutions in internet. It was necessary to spend money on hiring a data warehousing specialist to complete it..

Reference

SQLite Home Page. (n.d.). https://www.sqlite.org/index.html

Insausti, S. (2019, August 1). Running a Data Warehouse on PostgreSQL. Severalnines. https://severalnines.com/database-blog/running-data-warehouse-postgresql

Google. (n.d.). CSV file: Definition. Google Ads Help. https://support.google.com/google-ads/answer/9004364?hl=en

Bartolini, G. (2009, November 6). Data warehousing with PostgreSQL. 2ndQuadrant. https://wiki.postgresql.org/images/3/38/PGDay2009-EN-Datawarehousing_with_PostgreSQL.pdf

Martinez, N. (2021, January, 2021). Basketball Named 3rd Most Popular Sport In The World With Over 2 Billion Fans Worldwide. Fadeaway World. https://fadeawayworld.net/nba-media/basketball-named-3rd-most-popular-sport-in-the-world-with-over-2-billion-fans-worldwide

WorkoutHealthy. (2021, May 5). 5 Factors Influencing Sports Performance. https://www.workouthealthy.com/blog-sports-performance

https://www.geeksforgeeks.org/etl-process-in-data-warehouse/

https://www.ibm.com/docs/ko/cognos-analytics/10.2.2?topic=cubes-in-database-aggregates

https://blog.panoply.io/how-to-estimate-cloud-data-warehouse-costs-and-compare-pricing

APPENDIX 1

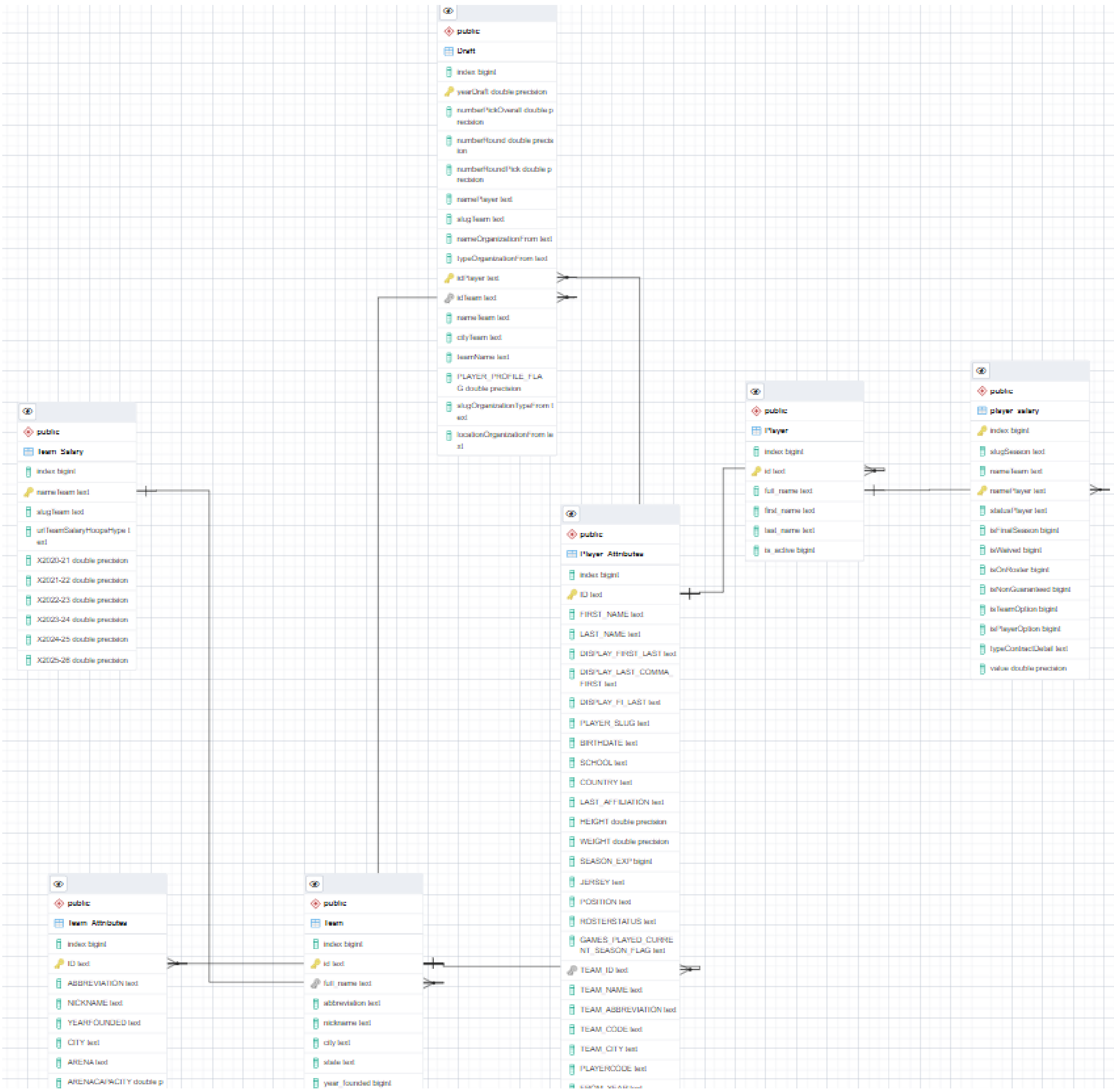
https://drive.google.com/file/d/1_SNj5Sxy8kq944xGsFQP-CDDbwl0Ci8R/view?usp=sharing

(It should be clearly seen through image in the google drive)

APPENDIX 2

https://drive.google.com/file/d/1AuvzQKgehleGMyTltLkhdHEgnvUW_Zy/view?usp=sharing

(It should see through image in the google drive)



APPENDIX 4

Query Editor

```
select AVG("Player_Attributes"."PTS"),"Player_Attributes"."FIRST_NAME", "Draft"."yearDraft"
From "Player_Attributes"
inner join "Draft"
on "Player_Attributes"."ID" = "Draft"."idPlayer"
group BY "Player_Attributes"."FIRST_NAME","Draft"."yearDraft"
order by "Draft"."yearDraft" asc
```

Data Output | Query History | Explain

a Output

	avg double precision	FIRST_NAME text	yearDraft double precision
1	22.5	Alex	1949
2	5.425	Bob	1949
3	3.7	Cliff	1949
4	8	Dick	1949
5	9	Don	1949
6	8.2	Duane	1949
7	1.7	Earl	1949
8	12.65	Ed	1949
9	9.5	Ernie	1949
10	5.6	Frank	1949

Messages

```
Successfully run. Total query runtime: 42 msec.
2939 rows affected.
```

APPENDIX 5

1. The data set were took from Kaggle and available in link below:
   https://www.kaggle.com/datasets/wyattowalsh/basketball

2. This the basket_dw backup file available in google drive link, may restore to run the database on PostgreSQL
   https://drive.google.com/file/d/1Qhr1y4xGM2gUneYeQxMQLxFNTqnOrHAI/view?usp=sharing
3. This is the link for the tableau .twbx file, may retrieved to view tableau OLAP and Visualizaton.
   https://drive.google.com/file/d/1Thex1d5_DhGyuf2zMbp-H1jl_2SnzHf_/view?usp=sharing
   https://drive.google.com/file/d/1sPJ5Ip0n2TQTIgps0e6eETK0FSyxe8c0/view?usp=sharing