In [1]:
```
#Install required packages (pandasql)
!pip install pandasql
```

```
Collecting pandasql
  Downloading pandasql-0.7.3.tar.gz (26 kB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'done'
  Preparing metadata (pyproject.toml): started
  Preparing metadata (pyproject.toml): finished with status 'done'
Requirement already satisfied: numpy in c:\users\swath\appdata\local\program
s\python\python310\lib\site-packages (from pandasql) (1.24.2)
Requirement already satisfied: pandas in c:\users\swath\appdata\local\progra
ms\python\python310\lib\site-packages (from pandasql) (2.0.0)
Requirement already satisfied: sqlalchemy in c:\users\swath\appdata\local\pr
ograms\python\python310\lib\site-packages (from pandasql) (2.0.9)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\swath\appd
ata\local\programs\python\python310\lib\site-packages (from pandas->pandasq
l) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\swath\appdata\local
\programs\python\python310\lib\site-packages (from pandas->pandasql) (2023.
3)
Requirement already satisfied: tzdata>=2022.1 in c:\users\swath\appdata\loca
l\programs\python\python310\lib\site-packages (from pandas->pandasql) (2023.
3)
Requirement already satisfied: typing-extensions>=4.2.0 in c:\users\swath\ap
pdata\local\programs\python\python310\lib\site-packages (from sqlalchemy->pa
ndasql) (4.5.0)
Requirement already satisfied: greenlet!=0.4.17 in c:\users\swath\appdata\lo
cal\programs\python\python310\lib\site-packages (from sqlalchemy->pandasql)
(2.0.2)
Requirement already satisfied: six>=1.5 in c:\users\swath\appdata\local\prog
rams\python\python310\lib\site-packages (from python-dateutil>=2.8.2->pandas
->pandasql) (1.16.0)
Building wheels for collected packages: pandasql
  Building wheel for pandasql (pyproject.toml): started
  Building wheel for pandasql (pyproject.toml): finished with status 'done'
  Created wheel for pandasql: filename=pandasql-0.7.3-py3-none-any.whl size=
26801 sha256=7d371f7b91e853354108688da9ed65dd088f4228a54a7293d7cb7301d26a018
5
  Stored in directory: c:\users\swath\appdata\local\pip\cache\wheels\e9\bc\3
a\8434bdcccf5779e72894a9b24fecbdcaf97940607eaf4bcdf9
Successfully built pandasql
Installing collected packages: pandasql
Successfully installed pandasql-0.7.3
```

In [2]:
```
import pandas as pd
from pandasql import sqldf
```

In [ ]:
```
#Open and view dataset
```

In [76]:
```
df = pd.read_csv("D:\Certificates\SQL Jupyter Notebook\Indian Cricket Auction

df.head()
```

Out[76]:

	Player	Price	Type	Cost In ? (CR.)	Cost IN $ (000	2021 Squad	Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	1950.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	1950.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	1300.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	1170.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	1040.0	KKR	Gujarat Titans

In [6]:
```
Base Price
```

Out[6]:

	Player	Base Price	Type	Cost In ? (CR.)	Cost IN $ (000	2021 Squad	Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	1950.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	1950.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	1300.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	1170.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	1040.0	KKR	Gujarat Titans

In [ ]:
```
#  Data Exploration
```

In [7]:
```
#Total number of rows and columns

print('Total number of Rows: ', df.shape[0])
print('Total number of Columns: ', df.shape[1])
```

```
Total number of Rows:  633
Total number of Columns:  7
```

In [8]:
```
#Data Types

df.dtypes
```

Out[8]:
```
Player              object
Base Price          object
Type                object
Cost In ? (CR.)    float64
Cost IN $ (000     float64
2021 Squad          object
Team                object
dtype: object
```

In [11]:
```
#Total number of Rows
df.shape[0]
```

Out[11]:
```
633
```

In [13]:
```
#Total number of Column
df.shape[1]
```

Out[13]:
```
7
```

In [23]:
```
#Retreive Team data

df['Team']
```

Out[23]:
```
0      Gujarat Titans
1      Gujarat Titans
2      Gujarat Titans
3      Gujarat Titans
4      Gujarat Titans
            ...
628           Unsold
629           Unsold
630           Unsold
631           Unsold
632           Unsold
Name: Team, Length: 633, dtype: object
```

In [25]:
```
df = pd.read_csv("D:\Certificates\SQL Jupyter Notebook\Indian Cricket Auction

df.head()
```

Out[25]:

	Player	Base_Price	Type	Cost In ? (CR.)	Cost IN $ (000	2021 Squad	Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	1950.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	1950.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	1300.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	1170.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	1040.0	KKR	Gujarat Titans

In [27]:
```
#Retreive types data

df['Type'].unique()
```

Out[27]:
```
array(['BOWLER', 'ALL-ROUNDER', 'BATTER', 'WICKETKEEPER'], dtype=object)
```

In [33]:
```
df = pd.read_csv("D:\Certificates\SQL Jupyter Notebook\Indian Cricket Auction

df.head()
```

Out[33]:

	Player	Price	Type	Cost In ? (CR.)	Cost IN $ (000	2021 Squad	Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	1950.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	1950.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	1300.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	1170.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	1040.0	KKR	Gujarat Titans

In [36]:
```
#Retreive all column data

print(df.columns)
```

```
Index(['Player ', 'Price ', 'Type', 'Cost In ? (CR.)', 'Cost IN $ (000',
       '2021 Squad', 'Team'],
      dtype='object')
```

In [37]:
```
#Retreive Price data

df['Price ']
```

Out[37]:
```
0      Draft Pick
1      Draft Pick
2            2 Cr
3         40 Lakh
4      Draft Pick
           ...
628       20 Lakh
629       20 Lakh
630       20 Lakh
631       20 Lakh
632       20 Lakh
Name: Price , Length: 633, dtype: object
```

In [38]:
```
#Unique Values in 'Base Price' Column

df['Price '].unique()
```

Out[38]:
```
array(['Draft Pick', '2 Cr', '40 Lakh', '20 Lakh', '1 Cr', '75 Lakh',
       '50 Lakh', '30 Lakh', 'Retained', '1.5 Cr'], dtype=object)
```

In [39]:
```
#Distribution of values inside column 'Base Price'

df['Price '].value_counts()
```

Out[39]:
```
Price
20 Lakh       344
50 Lakh       104
2 Cr           48
1 Cr           33
Retained       27
75 Lakh        26
1.5 Cr         20
40 Lakh        16
30 Lakh         9
Draft Pick      6
Name: count, dtype: int64
```

In [40]:
```
#Retreive type data

df['Type']
```

Out[40]:
```
0           BOWLER
1      ALL-ROUNDER
2           BOWLER
3      ALL-ROUNDER
4           BATTER
          ...
628         BATTER
629         BOWLER
630         BOWLER
631         BOWLER
632    ALL-ROUNDER
Name: Type, Length: 633, dtype: object
```

In [41]:
```
#Unique Values in 'TYPE' Column

df['Type'].unique()
```

Out[41]:
```
array(['BOWLER', 'ALL-ROUNDER', 'BATTER', 'WICKETKEEPER'], dtype=object)
```

In [42]:
```
#Distribution of values inside column 'Base Price'

df['Type'].value_counts(normalize=True)
```

Out[42]:
```
Type
ALL-ROUNDER     0.382306
BOWLER          0.339652
BATTER          0.176935
WICKETKEEPER    0.101106
Name: proportion, dtype: float64
```

In [ ]:
```
#Questions to Answer
                      #Top 3 batsman who got paid the most?
                      #Top 5 bowlers who got paid the most?
                      #Highest paid all-rounders?
                      #Average pay for Batsman, Bowler, All-Rounder, Wicket-Ke
                       #List of Retained players with Salary?
```

In [ ]:
```
                          #Data Transformation
```

In [43]:
```
#Rename columns and save it in the variable df2

df2=df.rename(columns={'Player ':'Players',
                       'Price ':'Base_Price',
                       'Type':'Types',
                       'Cost In ? (CR.)':'Cost_INR',
                       'Cost IN $ (000': 'Cost_USD',
                       '2021 Squad':'IPL_2021_Team',
                       'Team':'IPL_2022_Team'})
```

In [44]:
```
#Dropping USD Column

df3 = df2.drop(['Cost_USD'],axis=1)
```

In [45]:
```
#Check updated Dataframe

df3.head()
```

Out[45]:

	Players	Base_Price	Types	Cost_INR	IPL_2021_Team	IPL_2022_Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	KKR	Gujarat Titans

In [ ]:
```
#Setting up mysql function to run queries
#Basics : The main function used in pandasql is sqldf. sqldf accepts 2 parame

#Specifying locals() or globals() can get tedious. You can defined a short he
```

In [46]:
```
mysql = lambda q: sqldf(q, globals())
```

In [ ]:
```
                #Question 1 - Name top 3 batsman who got paid the most?
```

In [48]:
```
print(df.columns)
```

```
Index(['Player ', 'Price ', 'Type', 'Cost In ? (CR.)', 'Cost IN $ (000',
       '2021 Squad', 'Team'],
      dtype='object')
```

In [49]:
```
print(df3.columns)
```

```
Index(['Players', 'Base_Price', 'Types', 'Cost_INR', 'IPL_2021_Team',
       'IPL_2022_Team'],
      dtype='object')
```

In [52]:
```
mysql = lambda q: sqldf(q, globals())
```

In [58]:
```
mysql("""select * from df3""")
```

Out[58]:

	Players	Base_Price	Types	Cost_INR	IPL_2021_Team	IPL_2022_Team
0	Rashid Khan	Draft Pick	BOWLER	15.0	SRH	Gujarat Titans
1	Hardik Pandya	Draft Pick	ALL-ROUNDER	15.0	MI	Gujarat Titans
2	Lockie Ferguson	2 Cr	BOWLER	10.0	KKR	Gujarat Titans
3	Rahul Tewatia	40 Lakh	ALL-ROUNDER	9.0	RR	Gujarat Titans
4	Shubman Gill	Draft Pick	BATTER	8.0	KKR	Gujarat Titans
...	...	...	...	...	...	...
628	Sairaj Patil	20 Lakh	BATTER	NaN	None	Unsold
629	Monu Singh	20 Lakh	BOWLER	NaN	None	Unsold
630	Nivethan Radhakrishnan	20 Lakh	BOWLER	NaN	None	Unsold
631	Lance Morris	20 Lakh	BOWLER	NaN	None	Unsold
632	Aaron Hardie	20 Lakh	ALL-ROUNDER	NaN	None	Unsold

633 rows × 6 columns

In [63]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='BATTER' """)
```

Out[63]:

	Players	Cost_INR
0	Shubman Gill	8.0
1	David Miller	3.0
2	Abhinav Sadarangani	2.6
3	Jason Roy	2.0
4	Ruturaj Gaikwad	6.0
...	...	...
107	Henan Malik	NaN
108	Pukhraj Mann	NaN
109	Shashwat Rawat	NaN
110	Jake Weatherald	NaN
111	Sairaj Patil	NaN

112 rows × 2 columns

In [64]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='BATTER' ORDER BY 2 "
```

Out[64]:

	Players	Cost_INR
0	Suresh Raina	NaN
1	Steve Smith	NaN
2	Rajat Patidar	NaN
3	Aaron Finch	NaN
4	Marnus Labuschagne	NaN
...	...	...
107	Mayank Agarwal	12.00
108	Shreyas Iyer	12.25
109	Kane Williamson	14.00
110	Virat Kohli	15.00
111	Rohit Sharma	16.00

112 rows × 2 columns

In [ ]:
```
#Question 1 - Name top 3 batsman who got paid the most?
```

In [66]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='BATTER' ORDER BY 2 D
```

Out[66]:

	Players	Cost_INR
0	Rohit Sharma	16.0
1	Virat Kohli	15.0
2	Kane Williamson	14.0

In [ ]:
```
#Question 2 - Name top 5 bowlers who get paid the most?
    #2-Top 5 bowlers who got paid the most?
```

In [67]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='BOWLER' ORDER BY 2 D
```

Out[67]:

	Players	Cost_INR
0	Rashid Khan	15.00
1	Deepak Chahar	14.00
2	Jasprit Bumrah	12.00
3	Shardul Thakur	10.75
4	Lockie Ferguson	10.00

In [ ]:
```
```

In [68]:
```
mysql("""    """)
```

In [ ]:
```
                    #Question 3 - Name 5 lowest paid wicket-keeper?
```

In [69]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='WICKETKEEPER' """)
```

Out[69]:

	Players	Cost_INR
0	Matthew Wade	2.40
1	Wriddhiman Saha	1.90
2	MS Dhoni	12.00
3	Ambati Rayudu	6.75
4	N. Jagadeesan	0.20
...	...	...
59	Fazil Makaya	NaN
60	Ryan Rickelton	NaN
61	Sandeep Kumar Tomar	NaN
62	Siddhesh Wath	NaN
63	Hardik Tamore	NaN

64 rows × 2 columns

In [ ]:
```
                                #Question 3 - Name 5 lowest paid wicket-keeper?
```

In [73]:
```
mysql("""select Players, Cost_INR from df3 WHERE Types ='WICKETKEEPER' AND Co
```

Out[73]:

	Players	Cost_INR
0	N. Jagadeesan	0.2
1	Baba Indrajith	0.2
2	Jitesh Sharma	0.2
3	Aryan Juyal	0.2
4	Luvnith Sisodia	0.2

In [ ]:
```
                                #Question 4 - What is the Average pay for Batsman, Bowl
```

In [74]:
```
mysql("""SELECT Types, round(avg(Cost_INR),2) average_price FROM df3 GROUP BY
```

Out[74]:

	Types	average_price
0	WICKETKEEPER	5.09
1	BATTER	4.11
2	ALL-ROUNDER	3.61
3	BOWLER	3.07

In [ ]:
```
#Question 5 - List of Retained players with team name and
```

In [75]:
```
mysql("""SELECT PlayerS, Cost_INR FROM df3 WHERE Base_Price ="Retained" ORDER
```

Out[75]:

	Players	Cost_INR
0	Ravindra Jadeja	16.0
1	Rishabh Pant	16.0
2	Rohit Sharma	16.0
3	Virat Kohli	15.0
4	Sanju Samson	14.0

In [ ]:
```

```