

```
In [1]: ▶ # Import pandas
import pandas as pd
```

```
In [2]: ▶ # Set file path to a variable
input_file_path = r"C:\Users\kavis\Documents\Class-sem1\Data1202\Labs\vg-sales.csv"
print ("done")

done
```

```
In [3]: ▶ # Read Data into a DataFrame
input_raw_data = pd.read_csv(input_file_path)
```

```
In [4]: ▶ # How to create an engine and connect to a database

# import required libraries
import pandas as pd
from sqlalchemy import create_engine
import pymysql
print ("done")

done
```

```
In [5]: ▶ # create engine
engine = create_engine('mysql+pymysql://root:@127.0.0.1:51211')
print ("done")

done
```

```
In [6]: ▶ # connection string
conn = engine.connect()
print ("done")

done
```

```
In [7]: ▶ # read a simple query into DataFrame
df = pd.read_sql("SELECT * FROM data1202.vgsales", conn)
```

```
In [8]: ▶ # print DataFrame
df.head(10)
```

Out[8]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24
2	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82
3	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00
4	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37
5	6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26
6	7	New Super Mario Bros.	DS	2006	Platform	Nintendo	11.38	9.23	6.50	2.90	30.01
7	8	Wii Play	Wii	2006	Misc	Nintendo	14.03	9.20	2.93	2.85	29.02
8	9	New Super Mario Bros. Wii	Wii	2009	Platform	Nintendo	14.59	7.06	4.70	2.26	28.62
9	10	Duck Hunt	NES	1984	Shooter	Nintendo	26.93	0.63	0.28	0.47	28.31

```
In [9]: df_q1=pd.read_sql('''SELECT AVG(Global_sales),
CASE
WHEN year < 2005 THEN 'Pre-2005'
when year > 2005 THEN 'Post-2005'
ELSE '2005'
END AS Label
FROM data1202.vgsales
group by Label''', conn)
df_q1.head()
```

Out[9]:

	AVG(Global_sales)	Label
0	0.489448	Post-2005
1	0.649884	Pre-2005
2	0.488778	2005

```
In [10]: df_q2=pd.read_sql('''SELECT *,CASE WHEN year <2005 THEN 'pre-2005' ELSE 'post-2005'END AS Label FROM data1202.vgsales
df_q2.head()
```

Out[10]:

	Rank	Name	Platform	Year	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Label
0	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74	post-2005
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	pre-2005
2	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.82	post-2005
3	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00	post-2005
4	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.37	pre-2005

In [ ]:

