DBS Project task-2

In [1]:
```python
import numpy as np
import pandas as pd
```

In [2]:
```python
#Reading csv files into dataframe
category = pd.read_csv(r'/users/rohanchitte/downloads/data-2/category.csv')
date_tab = pd.read_csv(r'/users/rohanchitte/downloads/data-2/date_table.csv')
event = pd.read_csv(r'/users/rohanchitte/downloads/data-2/event.csv')
sales = pd.read_csv(r'/users/rohanchitte/downloads/data-2/sales.csv')
venue = pd.read_csv(r'/users/rohanchitte/downloads/data-2/venue.csv')
```

In [3]:
```python
print(venue.isnull().sum())
print(category.isnull().sum())
print(date_tab.isnull().sum())
print(event.isnull().sum())
print(sales.isnull().sum())
```

```
venueid         0
venuename       0
venuecity       0
venuestate      0
venueseats     15
dtype: int64
catid       0
catgroup    0
catname     0
catdesc     0
dtype: int64
dateid     0
caldate    0
day        0
week       0
month      0
qtr        0
year       0
holiday    0
dtype: int64
eventid      0
venueid      0
catid        0
dateid       0
eventname    0
starttime    0
dtype: int64
salesid       0
listid        0
sellerid      0
buyerid       0
eventid       0
dateid        0
qtysold       0
pricepaid     0
commission    0
saletime      0
dtype: int64
```

In [4]:
```python
# 130/200 values are 0, so we will replace null values with 0.
#Moreover, the objective of visualization does not deal with venueseats.
#So, data imputation becomes irrelevant.
venueseats = venue['venueseats']
count = 0
for i in venueseats:
    if i == 0:
        count = count + 1
print(count)  #130
```

```
130
```

In [5]:
```python
#Data Imputation
venue['venueseats'] = venue['venueseats'].fillna(0)
```

In [ ]:
```python
#Importing Data
```

In [6]:
```python
from sqlalchemy import create_engine
engine = create_engine('postgresql://:@localhost:5432/databaseproject')
```

In [7]:
```python
sales.to_sql('sales', engine)
venue.to_sql('venue', engine)
event.to_sql('event', engine)
date_tab.to_sql('date_table', engine)
category.to_sql('category', engine)
```

In [8]:
```python
#Running an sql query to test the import of the sales table
conn = engine.raw_connection()
cur = conn.cursor()
cur.execute('Select salesid FROM sales')
salesid = cur.fetchall()
len(salesid)
```

Out[8]:   172456