

```
In [1]: import pandas as pd
import os
from sqlalchemy import create_engine
import numpy as np
```

```
In [3]: #importing dataset
df = pd.read_csv('mov_audible_uncleaned.csv')
df.head(2)
```

Out[3]:

	name	author	narrator	time	releasedate	language	stars	price
0	Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	4/8/2008	English	5 out of 5 stars34 ratings	468
1	The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	1/5/2018	English	4.5 out of 5 stars41 ratings	820

```
In [4]: #Make the index column start from 1 instead of 0 because we might need to use this in SQL when removing duplicate
#the table
df.index = np.arange(1, len(df) + 1)
```

```
In [5]: df.head(2)
```

Out[5]:

	name	author	narrator	time	releasedate	language	stars	price
1	Geronimo Stilton #11 & #12	Writtenby:GeronimoStilton	Narratedby:BillLobely	2 hrs and 20 mins	4/8/2008	English	5 out of 5 stars34 ratings	468
2	The Burning Maze	Writtenby:RickRiordan	Narratedby:RobbieDaymond	13 hrs and 8 mins	1/5/2018	English	4.5 out of 5 stars41 ratings	820

```
In [6]: #Using environmental variables to hide database password
db_pass = os.environ.get('db_pass')
```

```
In [7]: #Creating engine using sqlalchemy create engine library
engine = create_engine(f'postgresql://postgres:{db_pass}@localhost:5432/movie_database')
```

```
In [8]: #Connecting to engine
con = engine.connect()
```

```
In [10]: #Migrating data to postgresql table for cleaning
df.to_sql('raw_movie_table', con)
```

```
In [ ]:
```

```
In [ ]:
```