First Name: Tuan Dat

Last Name: Tran

Import Libraries

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Import Data

In [2]:
```
bottle = pd.read_csv('bottle.csv', low_memory=False)
pd.set_option('display.float_format', lambda x:"%f"%x)
bottle.head()
```

Out[2]:

	Cst_Cnt	Btl_Cnt	Sta_ID	Date	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	...	R
0	1	1	054.0 056.0	1949-03	19-4903CR-HY-060-0930-05400560-0000A-3	0	10.500000	33.440000	nan	25.649000	...	
1	1	2	054.0 056.0	1949-03	19-4903CR-HY-060-0930-05400560-0008A-3	8	10.460000	33.440000	nan	25.656000	...	
2	1	3	054.0 056.0	1949-03	19-4903CR-HY-060-0930-05400560-0010A-7	10	10.460000	33.437000	nan	25.654000	...	
3	1	4	054.0 056.0	1949-03	19-4903CR-HY-060-0930-05400560-0019A-3	19	10.450000	33.420000	nan	25.643000	...	

	Cst_Cnt	Btl_Cnt	Sta_ID	Date	Depth_ID	Depthm	T_degC	Salnty	O2ml_L	STheta	...	R
4	1	5	054.0 056.0	1949-03	19-4903CR-HY-060-0930-05400560-0020A-7	20	10.450000	33.421000	nan	25.643000	...	

5 rows × 75 columns

Data management

In [3]:
```
bottle["Depthm"] = pd.to_numeric(bottle["Depthm"],errors="coerce")
bottle["Salnty"] = pd.to_numeric(bottle["Salnty"],errors="coerce")
bottle["R_PRES"] = pd.to_numeric(bottle["R_PRES"],errors="coerce")
bottle["O2ml_L"] = pd.to_numeric(bottle["O2ml_L"],errors="coerce")
bottle["T_degC"] = pd.to_numeric(bottle["T_degC"],errors="coerce")
```

In [4]:
```
bottle["Depthm"].describe()
```

Out[4]:
```
count   864863.000000
mean       226.831951
std        316.050259
min          0.000000
25%         46.000000
50%        125.000000
75%        300.000000
max       5351.000000
Name: Depthm, dtype: float64
```

In [5]:
```
bottle["Depthm"] = bottle["Depthm"].astype('category')
bottle["Depth_category"] = pd.cut(bottle["Depthm"], [0, 46, 125, 300, np.inf])
bottle["Depth_category"].describe()
```

Out[5]:
```
count            831453
unique                4
top       (46.0, 125.0]
freq             232803
Name: Depth_category, dtype: object
```

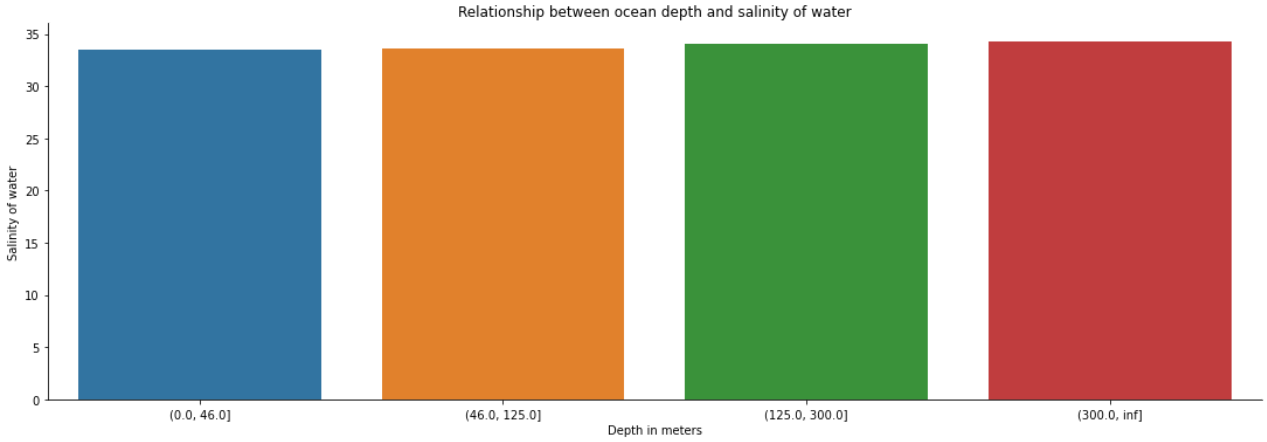
In [6]:
```
bottle["Salnty"].describe()
```

Out[6]:
```
count   817509.000000
mean        33.840350
std          0.461843
min         28.431000
25%         33.488000
50%         33.863000
75%         34.196900
max         37.034000
Name: Salnty, dtype: float64
```

Bar chart

In [7]:
```
%matplotlib inline
sns.catplot(x="Depth_category", y="Salnty", kind="bar", height=5, aspect=3, data=bottle
plt.title("Relationship between ocean depth and salinity of water")
plt.xlabel("Depth in meters")
plt.ylabel("Salinity of water")
```

Out[7]: Text(10.04999999999999, 0.5, 'Salinity of water')



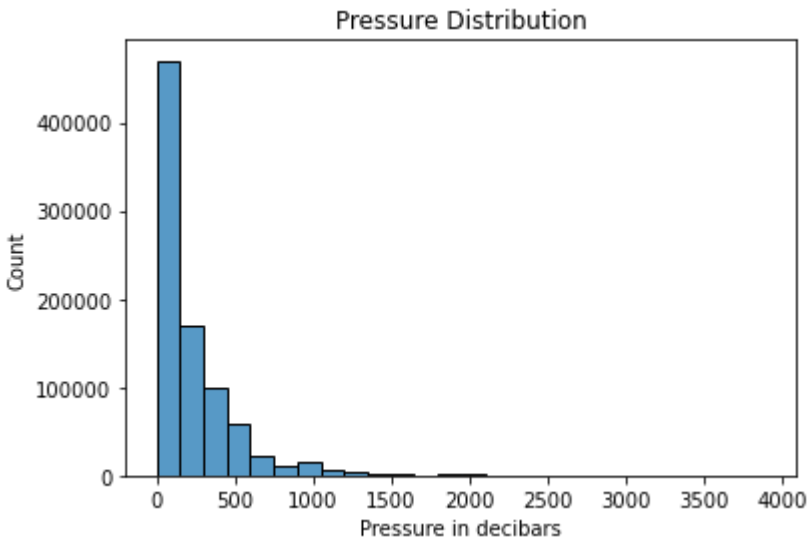
It is clear that the depth does not affect the salinity of water significantly.

Histogram

In [8]:
```
sns.histplot(bottle["R_PRES"].dropna(), bins=np.arange(0, 4000, 150))

plt.xlabel("Pressure in decibars")
plt.title("Pressure Distribution")
```

Out[8]: Text(0.5, 1.0, 'Pressure Distribution')



As can be seen from the histogram is that the pressure skews right. Therefore, the majority of water pressure are at 0dbar and the pressure higher, the number of pressure lower

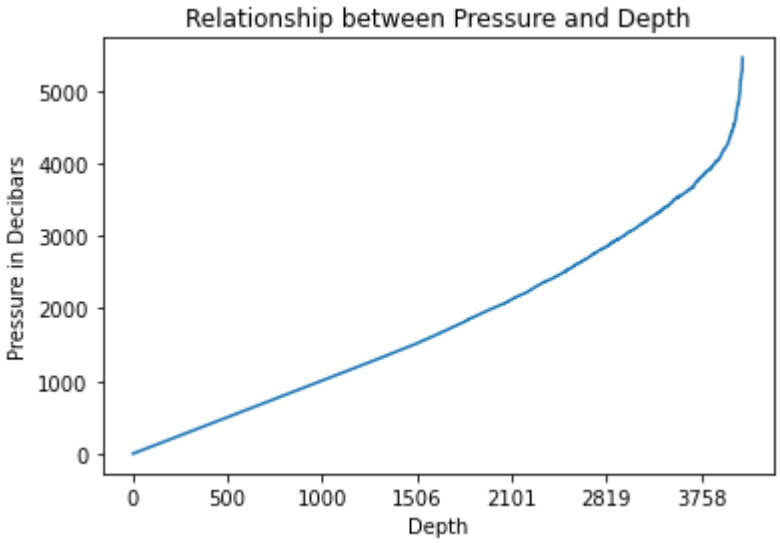
Line chart

In [9]:
```
var = bottle.groupby(["Depthm"]).R_PRES.mean()
print(var)
```

```
Depthm
0         0.000000
1         1.000672
2         2.000000
3         3.000000
4         4.002252
            ...
5154   5251.000000
5163   5260.000000
5165   5264.000000
5200   5302.000000
5351   5458.000000
Name: R_PRES, Length: 3219, dtype: float64
```

In [10]:
```
var.plot()
plt.title("Relationship between Pressure and Depth")
plt.ylabel("Pressure in Decibars")
plt.xlabel("Depth")
```

Out[10]:
```
Text(0.5, 0, 'Depth')
```



The line graph depicts the proportional increase of pressure to the depth of ocean remarkably.

Heat map

In [11]:
```
bottle["O2ml_L"].describe()
```

Out[11]:
```
count   696201.000000
mean         3.392468
std          2.073256
min         -0.010000
25%          1.360000
50%          3.440000
75%          5.500000
max         11.130000
Name: O2ml_L, dtype: float64
```

In [12]:
```
bottle["T_degC"].describe()
```

Out[12]:
```
count   853900.000000
```

```
mean        10.799677
std          4.243825
min          1.440000
25%          7.680000
50%         10.060000
75%         13.880000
max         31.140000
Name: T_degC, dtype: float64
```

In [13]:
```
# Divide water temperature into 4 categories

bottle["T_degC"] = bottle["T_degC"].astype('category')
bottle["T_degC_category"] = pd.cut(bottle["T_degC"], [1, 7, 10, 13, np.inf])
bottle.T_degC_category
```

Out[13]:
```
0         (10.0, 13.0]
1         (10.0, 13.0]
2         (10.0, 13.0]
3         (10.0, 13.0]
4         (10.0, 13.0]
              ...
864858     (13.0, inf]
864859     (13.0, inf]
864860     (13.0, inf]
864861     (13.0, inf]
864862     (13.0, inf]
Name: T_degC_category, Length: 864863, dtype: category
Categories (4, interval[float64]): [(1.0, 7.0] < (7.0, 10.0] < (10.0, 13.0] < (13.0, in
f]]
```

In [14]:
```
sub1 = bottle[["Depth_category","O2ml_L", "T_degC_category"]].copy()
sub1 = sub1.groupby(["Depth_category", "T_degC_category"]).mean().reset_index()
sub1
```

Out[14]:

	Depth_category	T_degC_category	O2ml_L
0	(0.0, 46.0]	(1.0, 7.0]	4.996667
1	(0.0, 46.0]	(7.0, 10.0]	4.584005
2	(0.0, 46.0]	(10.0, 13.0]	5.392275
3	(0.0, 46.0]	(13.0, inf]	5.660765
4	(46.0, 125.0]	(1.0, 7.0]	6.120294
5	(46.0, 125.0]	(7.0, 10.0]	3.378802
6	(46.0, 125.0]	(10.0, 13.0]	4.424456
7	(46.0, 125.0]	(13.0, inf]	5.243980
8	(125.0, 300.0]	(1.0, 7.0]	2.041272
9	(125.0, 300.0]	(7.0, 10.0]	2.387194
10	(125.0, 300.0]	(10.0, 13.0]	2.499583
11	(125.0, 300.0]	(13.0, inf]	3.504991
12	(300.0, inf]	(1.0, 7.0]	0.715749
13	(300.0, inf]	(7.0, 10.0]	0.849113
14	(300.0, inf]	(10.0, 13.0]	0.845561

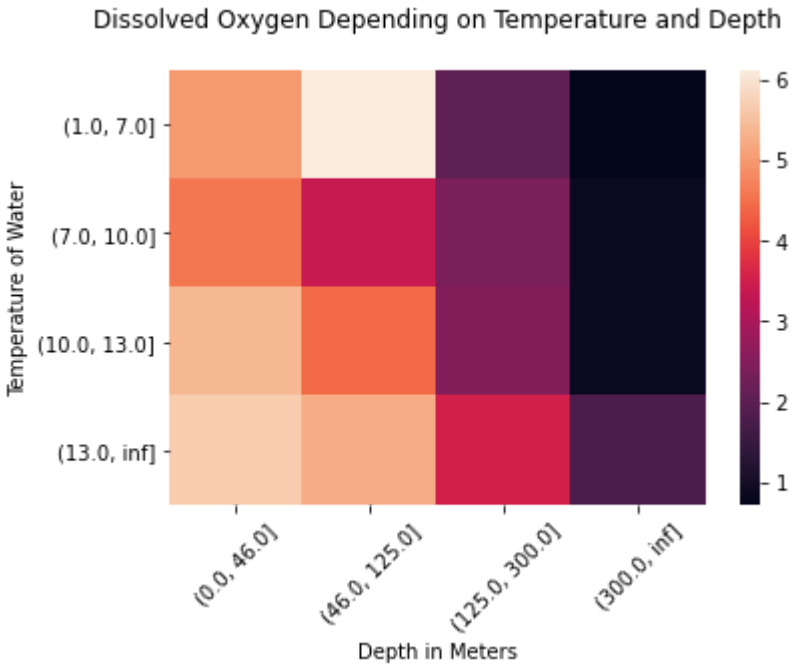
	Depth_category	T_degC_category	O2ml_L
15	(300.0, inf]	(13.0, inf]	1.806087

In [15]:
```
table = pd.pivot_table(data = sub1, index="T_degC_category", columns="Depth_category",
print(table)
```

```
Depth_category   (0.0, 46.0]  (46.0, 125.0]  (125.0, 300.0]  (300.0, inf]
T_degC_category
(1.0, 7.0]          4.996667       6.120294        2.041272      0.715749
(7.0, 10.0]         4.584005       3.378802        2.387194      0.849113
(10.0, 13.0]        5.392275       4.424456        2.499583      0.845561
(13.0, inf]         5.660765       5.243980        3.504991      1.806087
```

In [16]:
```
sns.heatmap(table)
plt.title("Dissolved Oxygen Depending on Temperature and Depth\n")
plt.xlabel("Depth in Meters")
plt.ylabel("Temperature of Water")
plt.xticks(rotation=45)
```

Out[16]:
```
(array([0.5, 1.5, 2.5, 3.5]),
 [Text(0.5, 0, '(0.0, 46.0]'),
  Text(1.5, 0, '(46.0, 125.0]'),
  Text(2.5, 0, '(125.0, 300.0]'),
  Text(3.5, 0, '(300.0, inf]')])
```



It is clear that the saturation of oxygen almost lowest in the deepest ocean layer whereas the highest amount of oxygen is in 46 to 125 meters with temperature from 1 to 7 degrees C