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

In [4]:
```
emp=pd.read_csv('C:\\Users\\DELL\\Downloads\\DS- Data Sets\\Simple linear regression\\emp_data.csv')
emp
```

Out[4]:

| | Salary_hike | Churn_out_rate |
|---|---|---|
| **0** | 1580 | 92 |
| **1** | 1600 | 85 |
| **2** | 1610 | 80 |
| **3** | 1640 | 75 |
| **4** | 1660 | 72 |
| **5** | 1690 | 70 |
| **6** | 1706 | 68 |
| **7** | 1730 | 65 |
| **8** | 1800 | 62 |
| **9** | 1870 | 60 |

In [5]:
```
emp.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Salary_hike     10 non-null     int64
 1   Churn_out_rate  10 non-null     int64
dtypes: int64(2)
memory usage: 288.0 bytes
```

In [6]:
```
emp.isnull()
```

Out[6]:

| | Salary_hike | Churn_out_rate |
|---|---|---|
| **0** | False | False |
| **1** | False | False |
| **2** | False | False |
| **3** | False | False |
| **4** | False | False |
| **5** | False | False |
| **6** | False | False |
| **7** | False | False |
| **8** | False | False |
| **9** | False | False |

In [7]:
```
emp.corr()
```

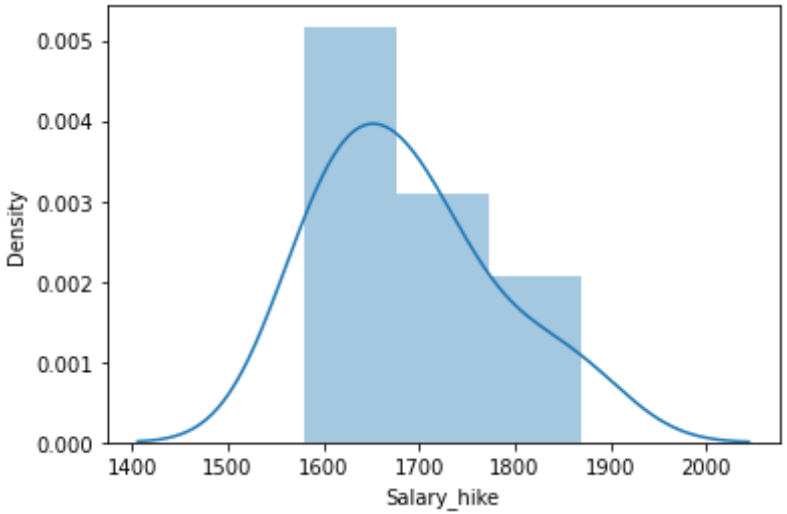
Out[7]:

| | Salary_hike | Churn_out_rate |
|---|---|---|
| **Salary_hike** | 1.000000 | -0.911722 |
| **Churn_out_rate** | -0.911722 | 1.000000 |

In [8]:
```
sns.distplot(emp['Salary_hike'])
```

```
C:\Users\DELL\anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated fu
nction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

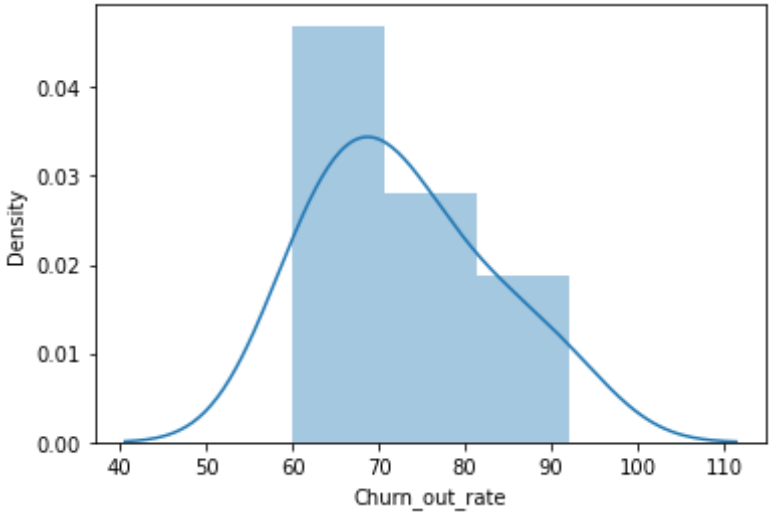
Out[8]:
```
<AxesSubplot:xlabel='Salary_hike', ylabel='Density'>
```




In [9]:
```
sns.distplot(emp['Churn_out_rate'])
```

```
C:\Users\DELL\anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated fu
nction and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level functi
on with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[9]: <AxesSubplot:xlabel='Churn_out_rate', ylabel='Density'>



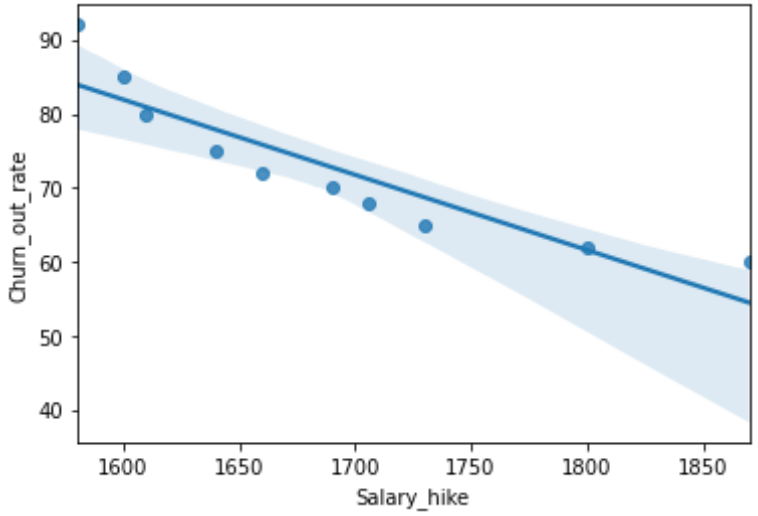
In [10]:
```
#model fitting
```

In [11]:
```
import statsmodels.formula.api as smf
```

In [12]:
```
model=smf.ols('Salary_hike~Churn_out_rate',data=emp).fit()
```

In [13]:
```
sns.regplot(x='Salary_hike',y='Churn_out_rate',data=emp)
```

Out[13]:
```
<AxesSubplot:xlabel='Salary_hike', ylabel='Churn_out_rate'>
```




In [14]:
```
#model parameters
model.params
```

Out[14]:
```
Intercept         2285.365297
Churn_out_rate      -8.186081
dtype: float64
```

In [15]:
```
#t and p values
model.tvalues ,'\n', model.pvalues
```

Out[15]:
```
(Intercept         23.827849
 Churn_out_rate    -6.277226
 dtype: float64,
```

```
'\n',
Intercept         1.024852e-08
Churn_out_rate    2.385777e-04
dtype: float64)
```

In [16]:
```
#rsquared and adj rsquared values
model.rsquared,model.rsquared_adj
```

Out[16]:
```
(0.8312363099883753, 0.8101408487369222)
```

In [17]:
```
model.summary()
```

```
C:\Users\DELL\anaconda3\lib\site-packages\scipy\stats\stats.py:1603: UserWarning: kurtosistest only valid for n>=20
... continuing anyway, n=10
  warnings.warn("kurtosistest only valid for n>=20 ... continuing "
```

Out[17]:

OLS Regression Results

| | | | |
|---|---|---|---|
| **Dep. Variable:** | Salary_hike | **R-squared:** | 0.831 |
| **Model:** | OLS | **Adj. R-squared:** | 0.810 |
| **Method:** | Least Squares | **F-statistic:** | 39.40 |
| **Date:** | Fri, 21 May 2021 | **Prob (F-statistic):** | 0.000239 |
| **Time:** | 21:51:10 | **Log-Likelihood:** | -49.995 |
| **No. Observations:** | 10 | **AIC:** | 104.0 |
| **Df Residuals:** | 8 | **BIC:** | 104.6 |
| **Df Model:** | 1 | | |
| **Covariance Type:** | nonrobust | | |

| | **coef** | **std err** | **t** | **P>\|t\|** | **[0.025** | **0.975]** |
|---|---|---|---|---|---|---|
| **Intercept** | 2285.3653 | 95.912 | 23.828 | 0.000 | 2064.193 | 2506.538 |
| **Churn_out_rate** | -8.1861 | 1.304 | -6.277 | 0.000 | -11.193 | -5.179 |

| | | | |
|---|---|---|---|
| **Omnibus:** | 2.758 | **Durbin-Watson:** | 0.591 |
| **Prob(Omnibus):** | 0.252 | **Jarque-Bera (JB):** | 1.564 |
| **Skew:** | 0.940 | **Prob(JB):** | 0.458 |

| | | | |
|---|---|---|---|
| **Kurtosis:** | 2.536 | **Cond. No.** | 556. |

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [ ]: