Do

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

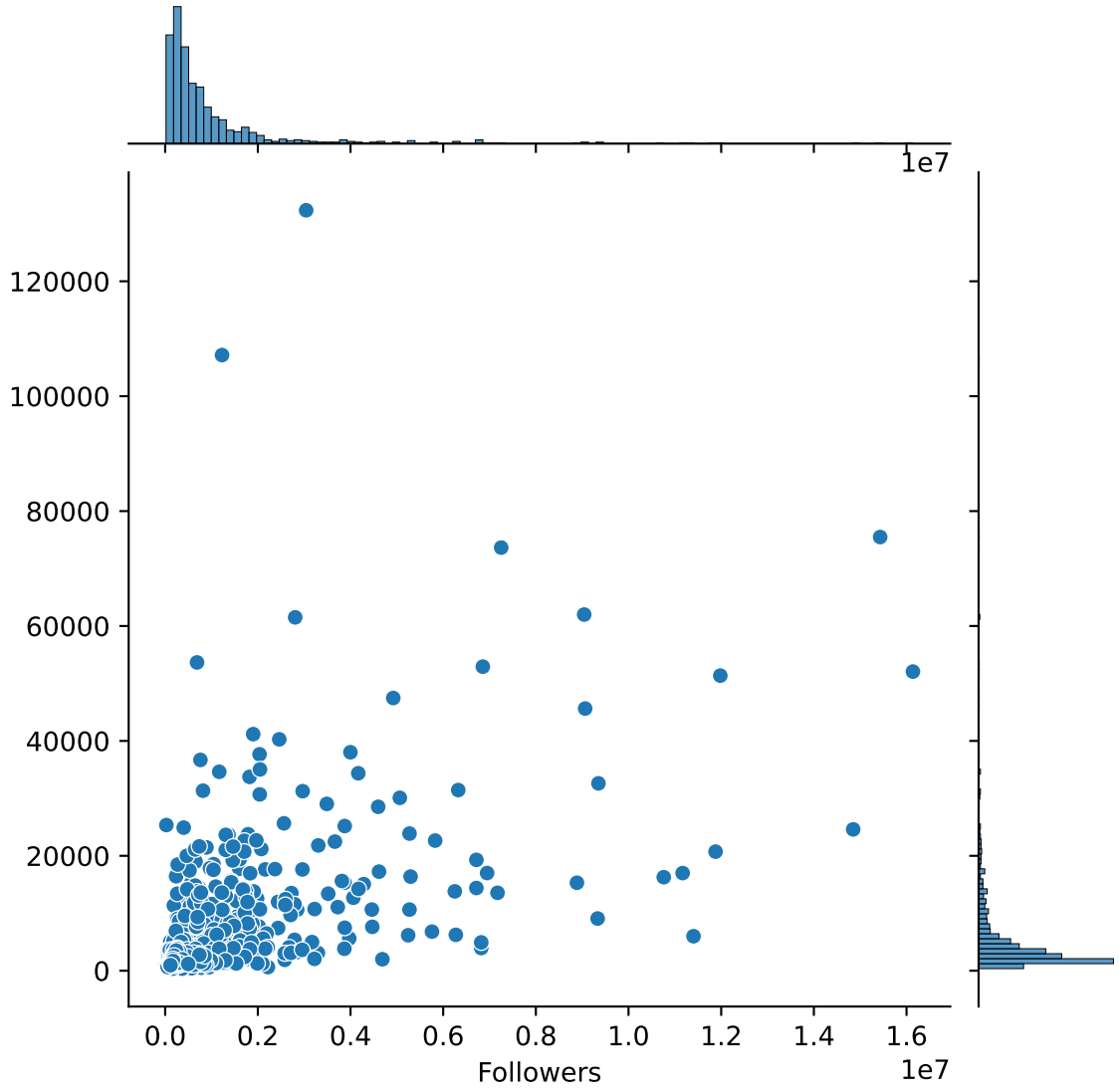
```
df = pd.read_csv("../Twitch_Streamer_Data_2023.csv")
df = pd.DataFrame(df)
```

```
# xs = np.arange(eng['Stream time'].min(), eng['Stream time'].max())
# ys = inverse(xs)
```

```
sns.jointplot(data=df, x='Followers', y='Average viewers')
```

```
## <seaborn.axisgrid.JointGrid object at 0x00000161C745D370>
```

```
plt.show()
```




```
# plt.plot(xs, ys, color = 'orange')
```

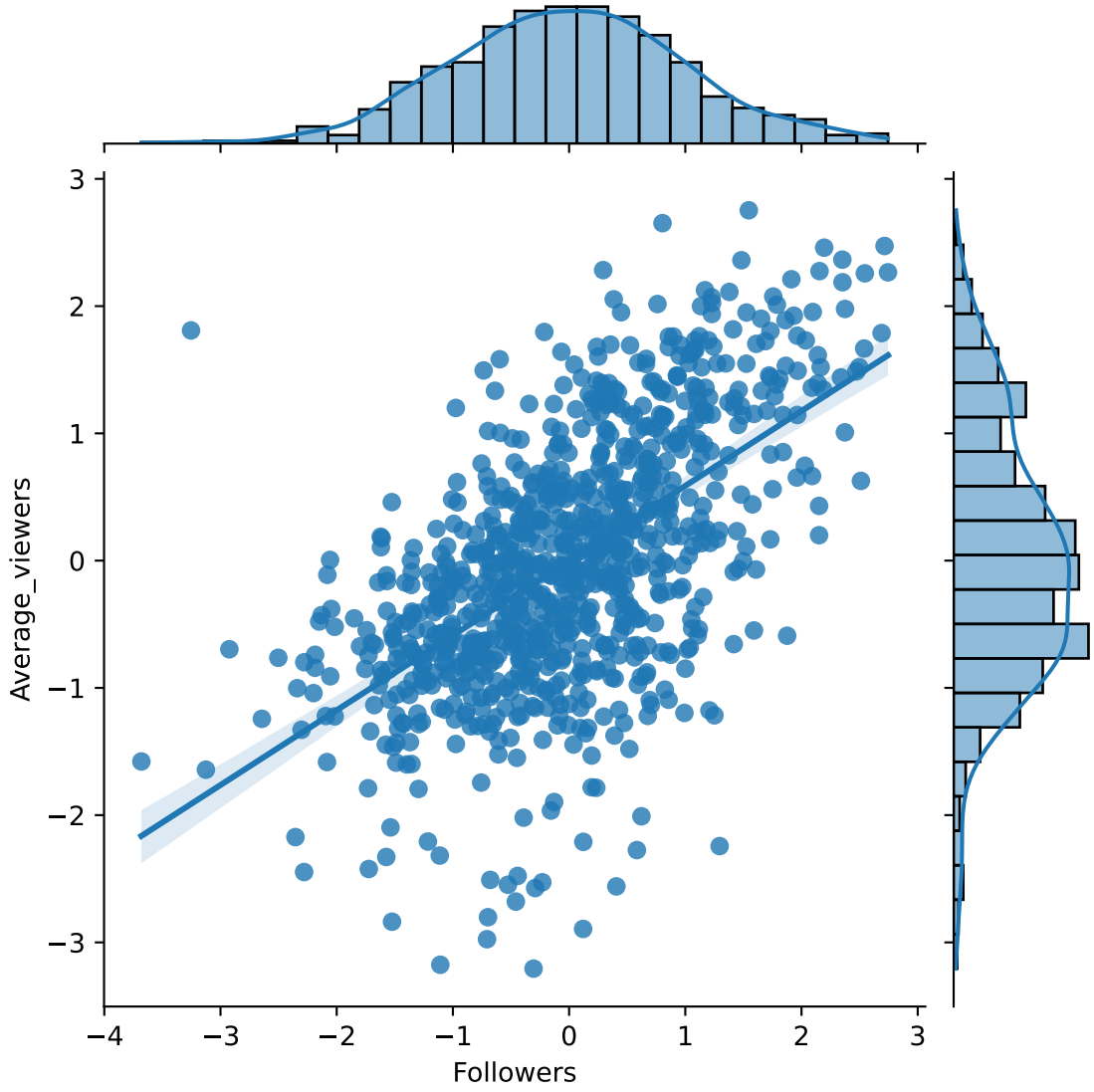
```
df.rename(columns={'Average viewers': 'Average_viewers', 'Mean weekly stream hours': 'Mean_weekly_strea
```

```
from sklearn.preprocessing import PowerTransformer
pt = PowerTransformer()
transformed = pt.fit_transform(df[['Followers', 'Average_viewers']])
df1 = pd.DataFrame(transformed)
df1.rename(columns={0: 'Followers', 1: 'Average_viewers'}, inplace = True)
```

```
sns.jointplot(data=df1, x = 'Followers', y ='Average_viewers', kind="reg")
```

```
## <seaborn.axisgrid.JointGrid object at 0x00000161CE990260>
```

```
plt.show()
```




```
import statsmodels.formula.api as smf
```

```
fit = smf.ols("Average_viewers ~ Mean_weekly_stream_hours", data=df).fit()
print(fit.summary())
```

```
##                             OLS Regression Results
## ==============================================================================
## Dep. Variable:        Average_viewers   R-squared:                       0.076
## Model:                            OLS   Adj. R-squared:                  0.075
## Method:                 Least Squares   F-statistic:                     73.49
## Date:                Fri, 12 Apr 2024   Prob (F-statistic):           4.39e-17
## Time:                        00:48:52   Log-Likelihood:                -9527.9
## No. Observations:                 900   AIC:                         1.906e+04
## Df Residuals:                     898   BIC:                         1.907e+04
## Df Model:                           1
## Covariance Type:            nonrobust
## ==============================================================================
##                                 coef    std err          t      P>|t|      [0.025      0.975]
```

```
## ------------------------------------------------------------------------------------------
## Intercept                     1e+04    538.353     18.580      0.000    8946.071    1.11e+04
## Mean_weekly_stream_hours   -94.4537     11.018     -8.572      0.000    -116.079     -72.829
## ==========================================================================
## Omnibus:                     1047.638   Durbin-Watson:                   1.159
## Prob(Omnibus):                  0.000   Jarque-Bera (JB):           100947.838
## Skew:                           5.739   Prob(JB):                         0.00
## Kurtosis:                      53.598   Cond. No.                         82.3
## ==========================================================================
##
## Notes:
## [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
print(df.dtypes)
```

```
## Channel_ID                           int64
## Channel                             object
## Watch time                           int64
## Stream time                          int64
## Peak viewers                         int64
## Average_viewers                      int64
## Followers Prev Yr                    int64
## Followers                            int64
## Followers gained                     int64
## Followers gained percent           float64
## Language                            object
## Partnered                            int64
## Mature                               int64
## Mean weekly watch hours            float64
## Mean_weekly_stream_hours           float64
## FollowersGainedPercentOutlier        int64
## FollowersGainedOutlier               int64
## WatchOutlier                         int64
## StreamOutlier                        int64
## PeakViewersOutlier                   int64
## FollowersOutlier                     int64
## AvgViewersOutlier                    int64
## Include                              int64
## Bin(Mean weekly stream hours)       object
## Bin(Mean weekly stream hours).1     object
## Bin(Mean weekly stream hours).2     object
## Median_avg_viewers                 float64
## Sequence                           float64
## Sequence.1                         float64
## Bin(Mean weekly stream hours).3     object
## Mean                               float64
## Median_stream_hrs                  float64
## dtype: object
```

```
model = smf.logit(formula='Partnered ~ Mean_weekly_stream_hours', data=df).fit()
```

```
## Optimization terminated successfully.
##          Current function value: 0.134643
##          Iterations 8
```

```
print(model.summary())
```

```
##                            Logit Regression Results
## ==============================================================================
## Dep. Variable:              Partnered   No. Observations:                  900
## Model:                          Logit   Df Residuals:                      898
## Method:                           MLE   Df Model:                            1
## Date:                Fri, 12 Apr 2024   Pseudo R-squ.:               0.0007377
## Time:                        00:48:53   Log-Likelihood:                -121.18
## converged:                       True   LL-Null:                       -121.27
## Covariance Type:            nonrobust   LLR p-value:                    0.6723
## ==============================================================================
##                                 coef    std err          z      P>|z|      [0.025      0.975]
## ------------------------------------------------------------------------------
## Intercept                     3.3592      0.341      9.864      0.000       2.692       4.027
## Mean_weekly_stream_hours      0.0031      0.008      0.408      0.683      -0.012       0.018
## ==============================================================================
```