HOW ARE THE PRICES OF VARIOUS NETFLIX SUBSCRIPTIONS DETERMINED IN A COUNTRY?

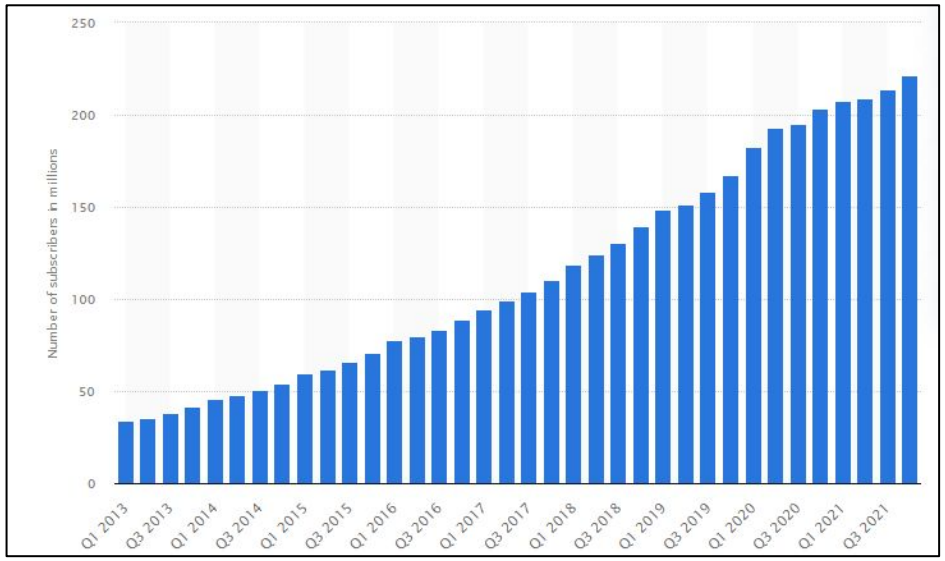
SC1015 MINI PROJECT

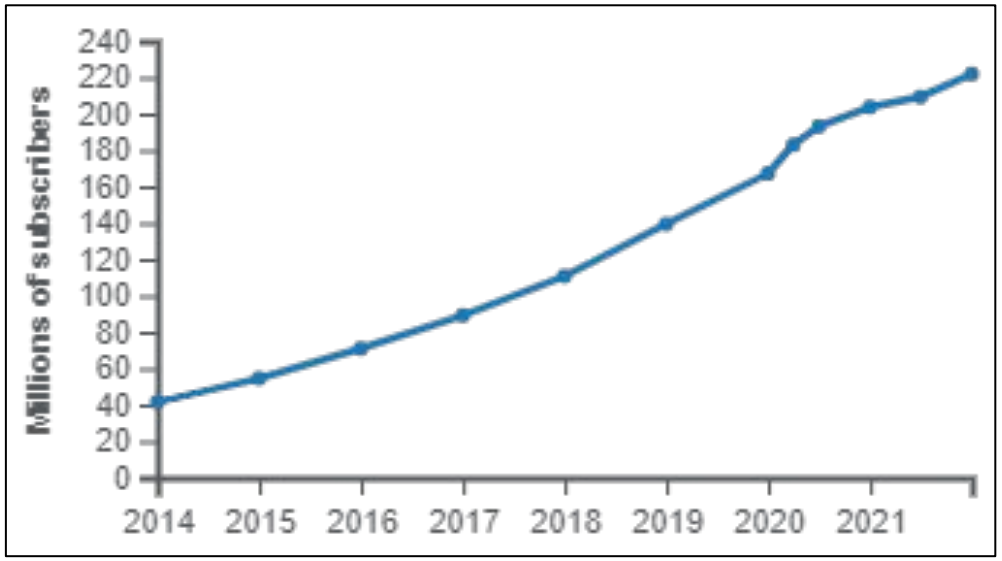
TERENCE OH
DONNA CHUA
KIAN HWEE

NETFLIX

UNDERSTANDING NETFLIX

NETFLIX USER BASE





- January 2022, 222 million subscribers worldwide
- Annual revenue of $29 Billion USD in 2021

NETFLIX

MOTIVATORS

Businesses operating internationally

Why exactly are prices almost always different in an international market?

NETFLIX

MOTIVATORS

Country	Total Library Size	Library Size Change (Aug 18 to Apr 22)	No. of TV Shows	No. of Movies	Cost Per Month - Basic ($)	Cost Per Month - Standard ($)	Cost Per Month - Premium ($)
Algeria	4,868		1,830	3,038	7.99	9.99	11.99
Argentina	5,049		1,900	3,149	3.38	5.70	8.37
Malaysia	6,025		2,445	3,580	8.29	10.66	13.03
Austria	5,794		1,916	3,878	8.09	14.13	19.37
Singapore	6,427		2,279	4,148	9.52	12.82	16.12
Belgium	5,095		1,674	3,421	9.78	14.67	19.57
Bermuda	6,273		2,103	4,170	8.99	12.99	15.99
Bolivia	5,042		1,895	3,147	7.99	10.99	13.99
Brazil	5,016		1,865	3,151	5.51	8.50	11.90
Bulgaria	7,162		2,024	5,138	8.69	10.87	13.04

Netflix in different countries

Experience of netflix in every country is different with different shows, prices & movies

NETFLIX

SOME POSSIBLE FACTORS

- Purchasing power/Income

- Population Density

- Library Size

NETFLIX

DATASETS

https://www.kaggle.com/datasets/prasertk/netflix-subscription-price-in-different-countries?select=netflix+price+in+different+countries.csv

kaggle

https://www.kaggle.com/datasets/tanuprabhu/population-by-country-2020

https://www.kaggle.com/datasets/nitishabharathi/gdp-per-capita-all-countries

NETFLIX

DATA CLEANING

DATA CLEANING

```
netflixdf = pd.read_csv('netflixunclean.csv')
incomedf = pd.read_csv('gdp.csv')
populationdf = pd.read_csv('population.csv')
```

```
df_merged = pd.merge(netflixdf, populationdf, on=['Country'], how='left')
fulldf = pd.merge(df_merged, incomedf, on=['Country'], how='left')
fulldf.count()
```

```
fulldf.columns
```

```
Index(['Country', 'Total Library Size', 'No. of TV Shows', 'No. of Movies',
       'Cost Per Month - Basic ($)', 'Cost Per Month - Standard ($)',
       'Cost Per Month - Premium ($)', 'Population (2020)', 'Yearly Change',
       'Net Change', 'Density (P/Km²)', 'Land Area (Km²)', 'Migrants (net)',
       'Fert. Rate', 'Med. Age', 'Urban Pop %', 'World Share', 'Country Code',
       '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
       '1999', '2000', '2001', '2002', '2003', '2004', '2005', '2006', '2007',
       '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016',
       '2017', '2018', '2019'],
      dtype='object')
```

```
fulldf.drop(['Land Area (Km²)', 'Migrants (net)',
       'Fert. Rate', 'Med. Age', 'Urban Pop %', 'World Share', 'Country Code',
       '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
       '1999', '2000', '2001', '2002', '2003', '2004', '2005', '2006', '2007',
       '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016',
       '2017', '2019', 'Population (2020)', 'Yearly Change',
       'Net Change', 'No. of TV Shows', 'No. of Movies'], axis=1, inplace=True)
```

```
fulldf = fulldf.dropna()
fulldf = fulldf[fulldf['2018'] >= 0]
fulldf = fulldf[fulldf['Density (P/Km²)'] >= 0]
fulldf = fulldf[fulldf['Cost Per Month - Basic ($)'] >= 0]
fulldf = fulldf[fulldf['Cost Per Month - Standard ($)'] >= 0]
fulldf = fulldf[fulldf['Cost Per Month - Premium ($)'] >= 0]
fulldf = fulldf[fulldf['Total Library Size'] >= 0]
```

```
fulldf.to_csv("C:\\Users\\teren\\Desktop\\cleaned_data.csv")
```

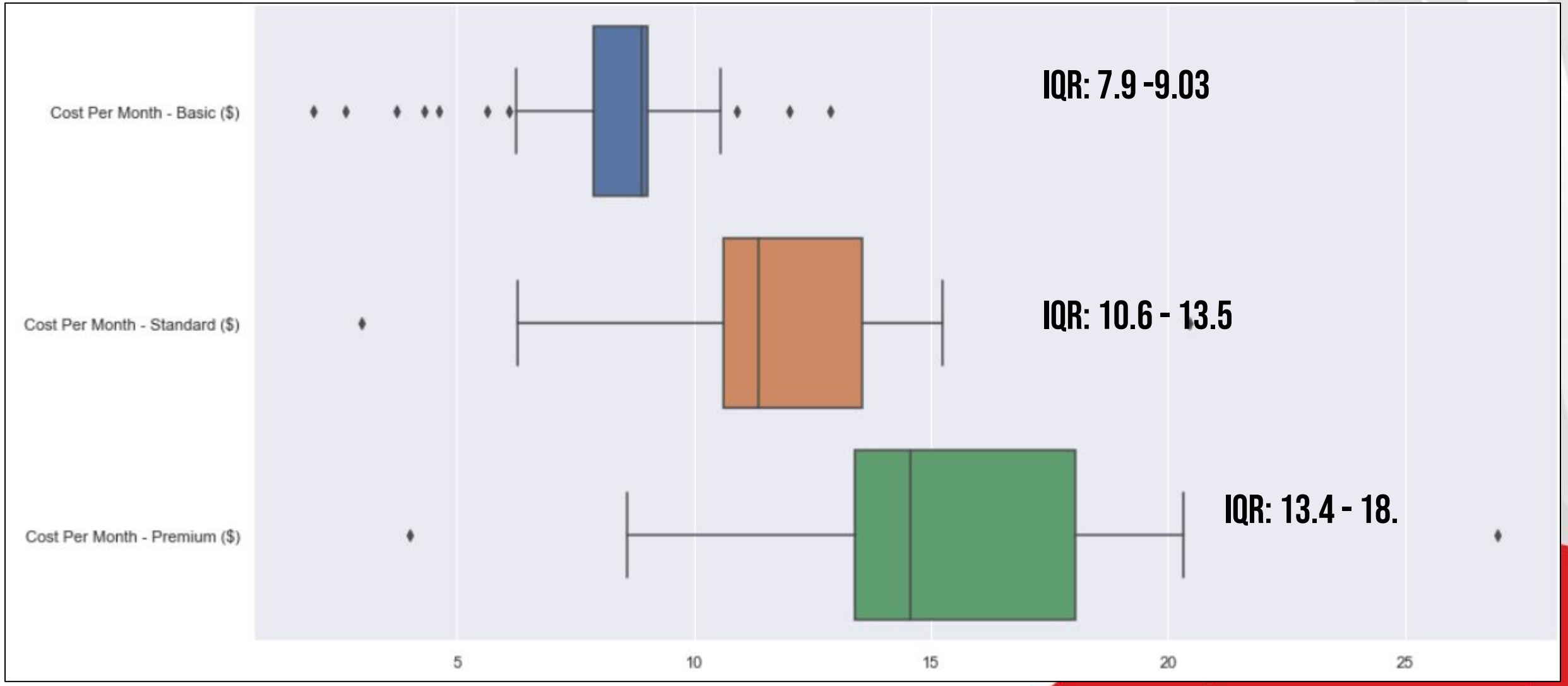
DATA CLEANING

Kept	Removed
• Country • Total Library Size • Cost per month - Basic($) • Cost per month - Standard($) • Cost per month - Premium($) • Density (P/km^2) //Population density • 2018 //Income	• 1990~2017 //Income • Land Area • Migrants • Fertility Rate • Age • Urban Population% • World Share • Country Code • Population • Yearly Change • Net Change • No. of TV Shows //Contributed to Library Size • No. of Movies //Contributed to Library Size

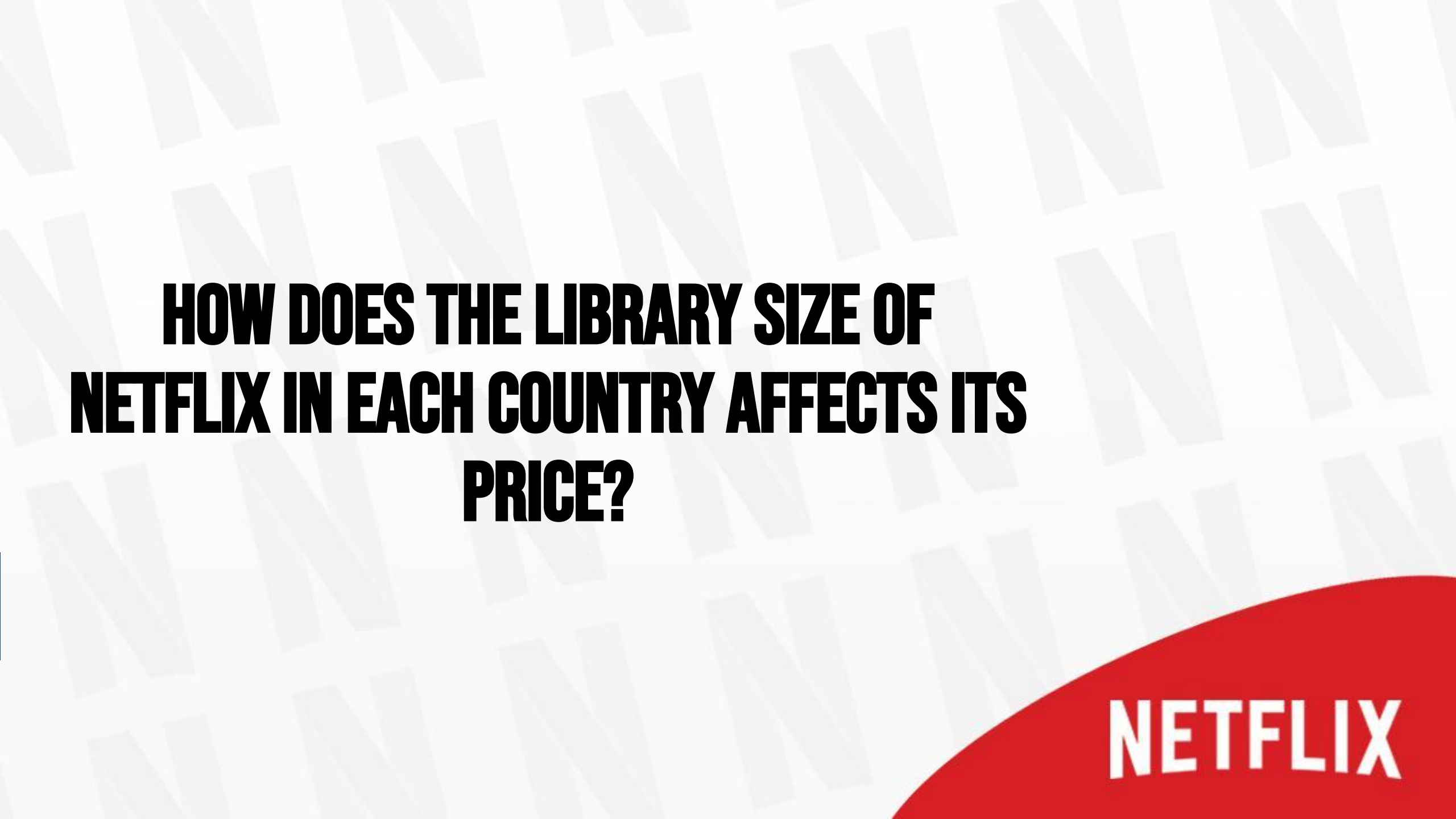
NETFLIX

DATA EXPLORATION

LOOKING AT OUR DATA



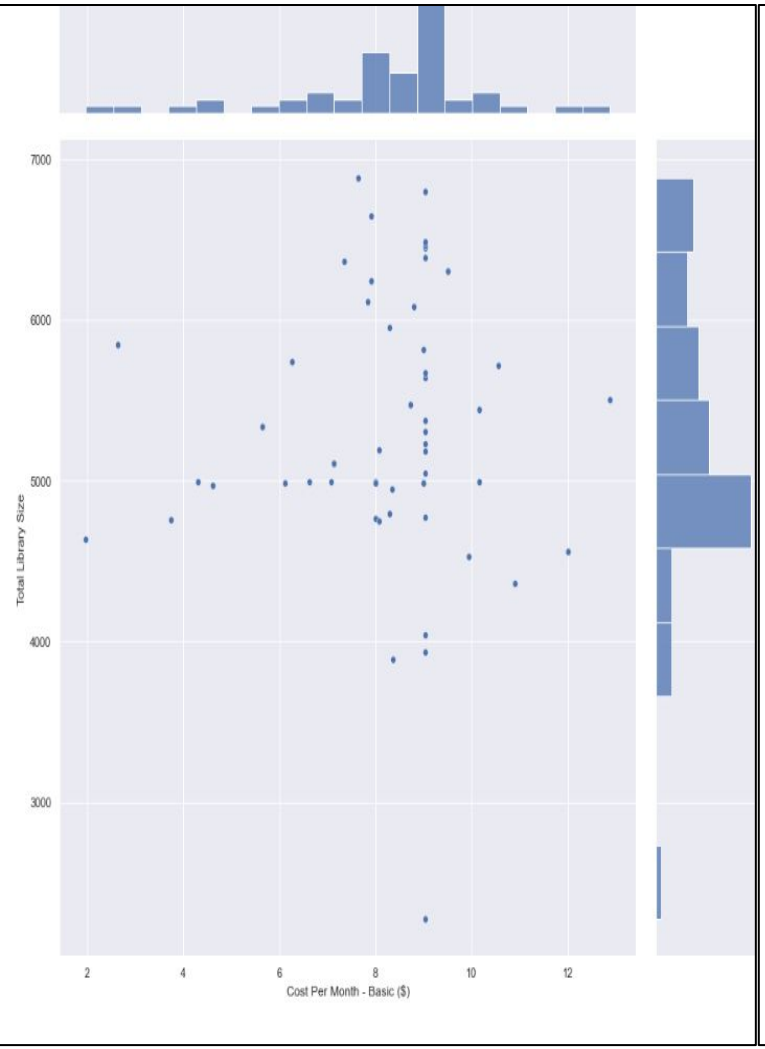
HOW DOES THE LIBRARY SIZE OF NETFLIX IN EACH COUNTRY AFFECTS ITS PRICE?

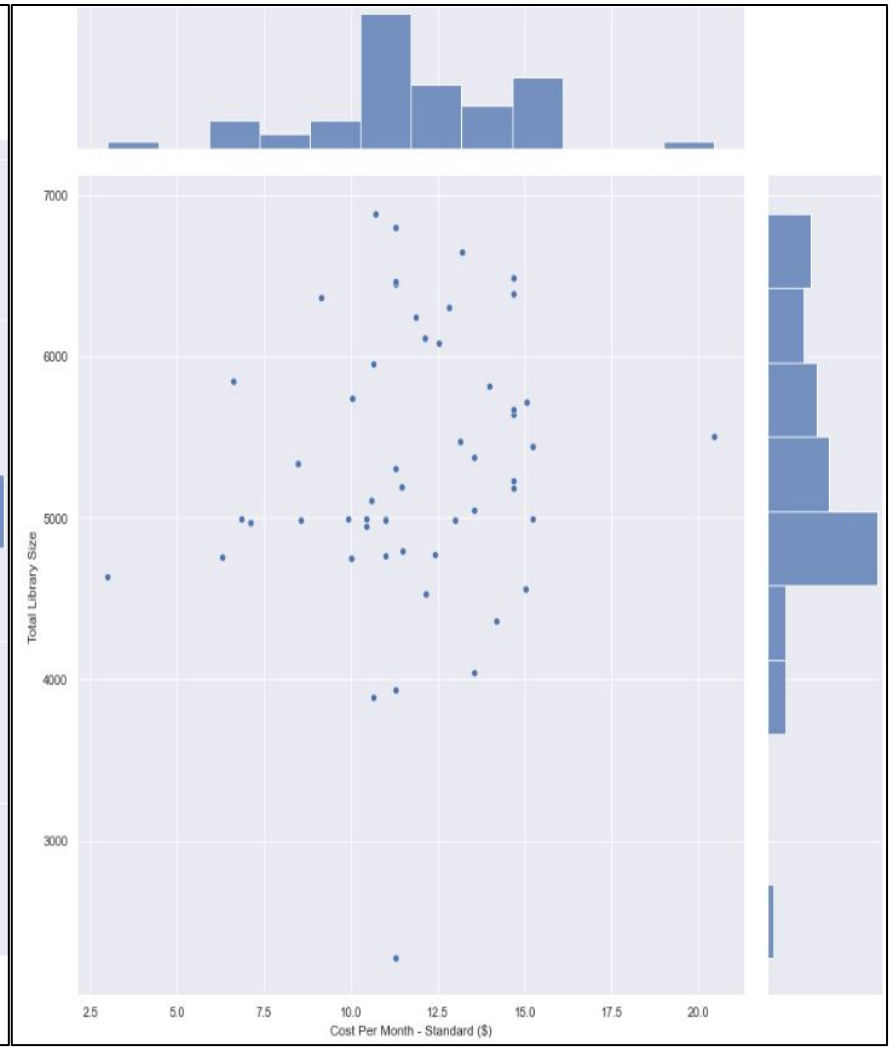


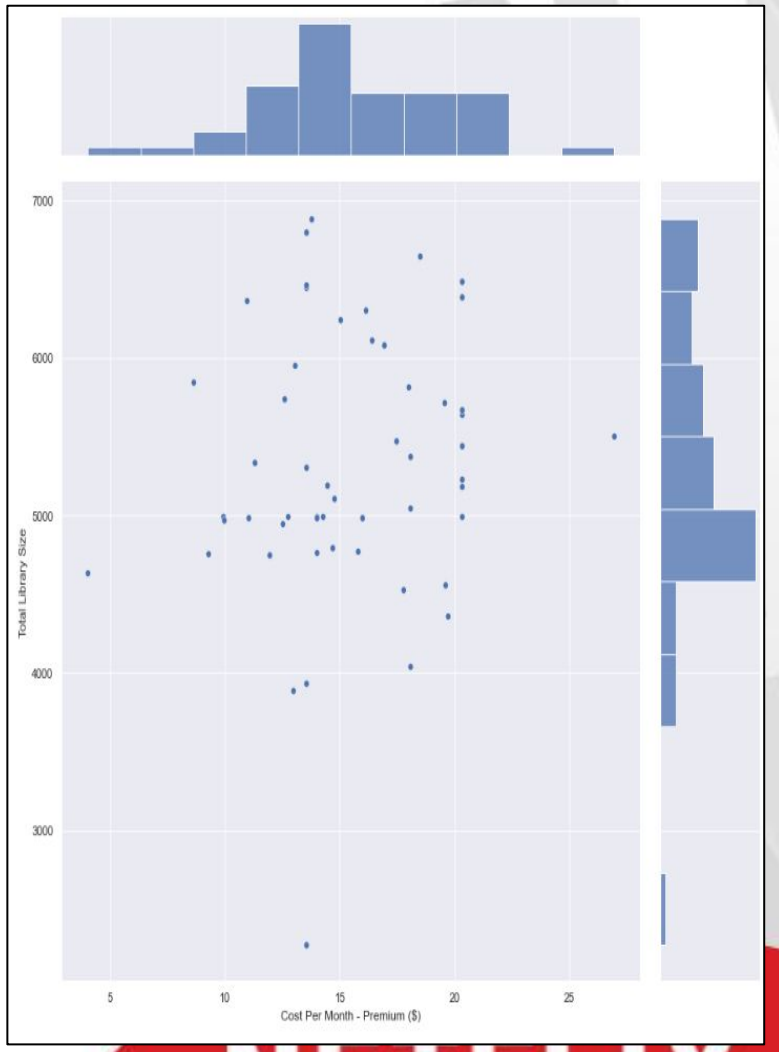
Cost Per Month - Basic ($)	Total Library Size
1.000000	0.020559
0.020559	1.000000

Cost Per Month - Standard ($)	Total Library Size
1.000000	0.127671
0.127671	1.000000

Cost Per Month - Premium ($)	Total Library Size
1.000000	0.125713
0.125713	1.000000



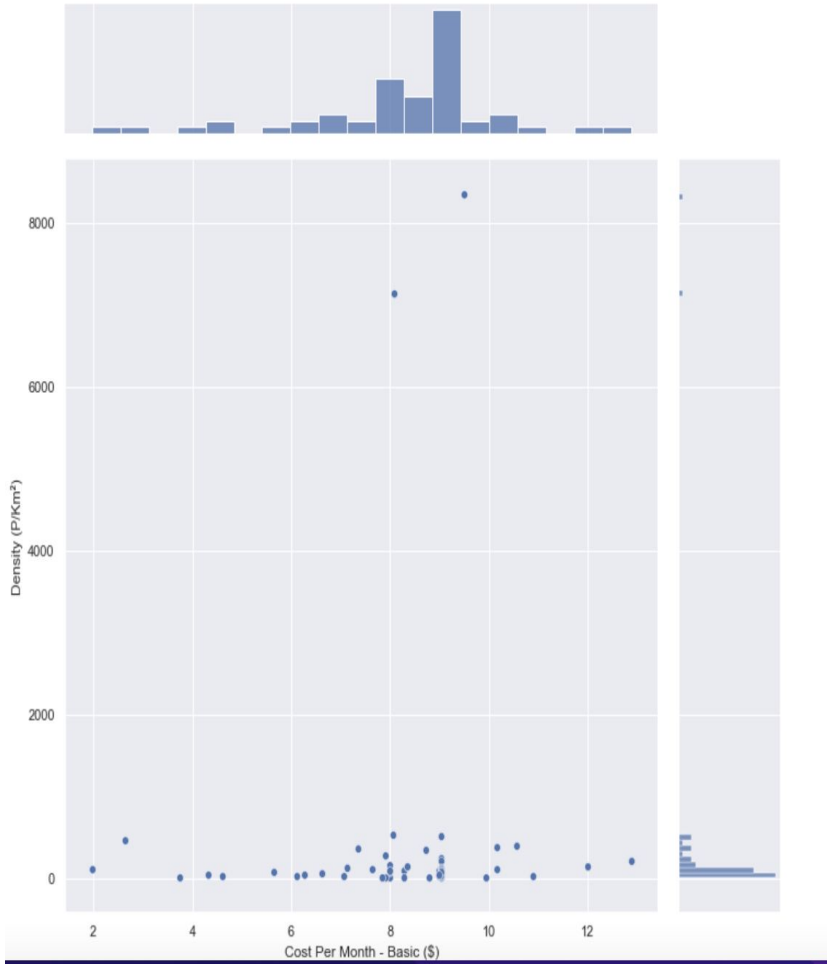




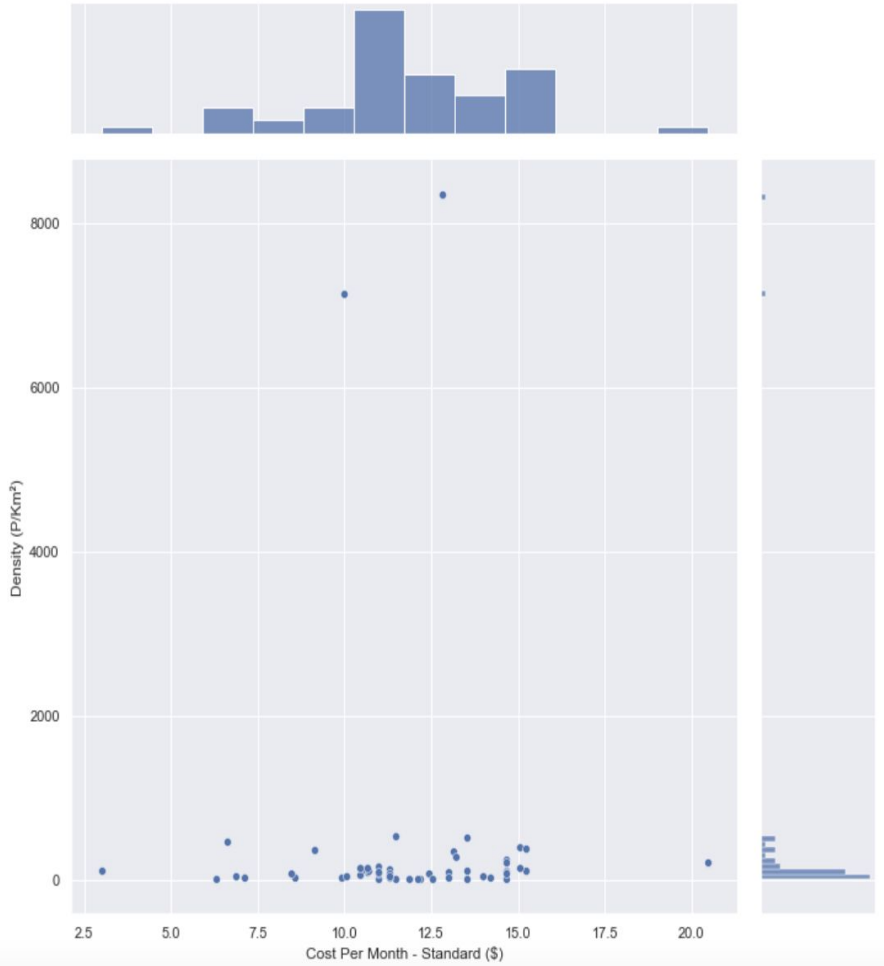
BASIC STANDARD PREMIUM

DOES A COUNTRY'S POPULATION DENSITY AFFECT THE PRICE OF NETFLIX SUBSCRIPTION?

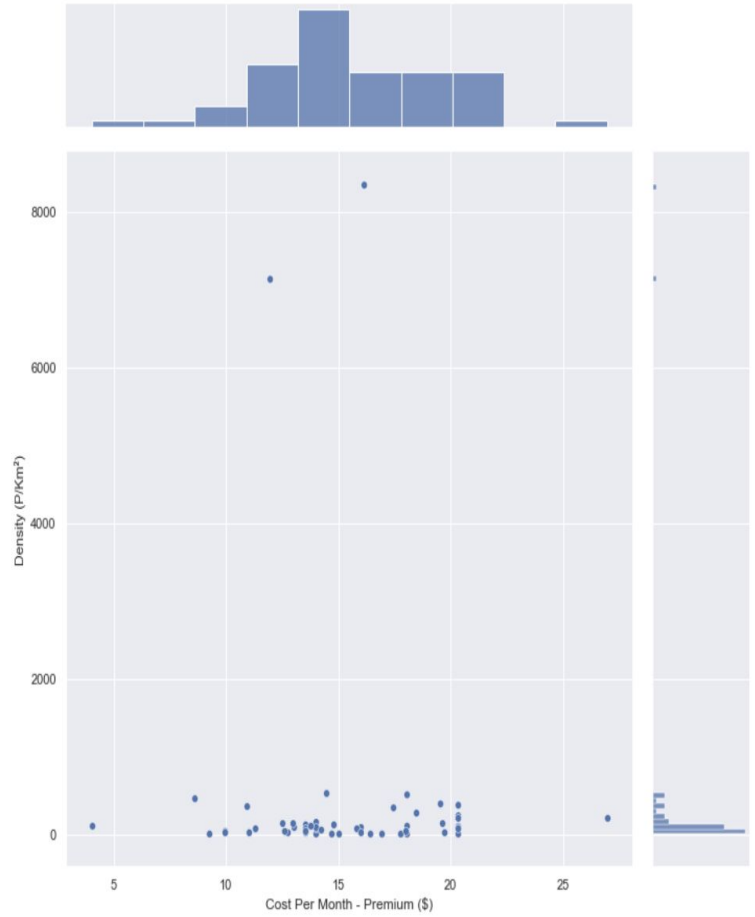
	Cost Per Month - Basic ($)	Density (P/Km²)
Cost Per Month - Basic ($)	1.000000	0.066314
Density (P/Km²)	0.066314	1.000000



	Cost Per Month - Standard ($)	Density (P/Km²)
Cost Per Month - Standard ($)	1.000000	-0.004108
Density (P/Km²)	-0.004108	1.000000



	Cost Per Month - Premium ($)	Density (P/Km²)
Cost Per Month - Premium ($)	1.000000	-0.044225
Density (P/Km²)	-0.044225	1.000000



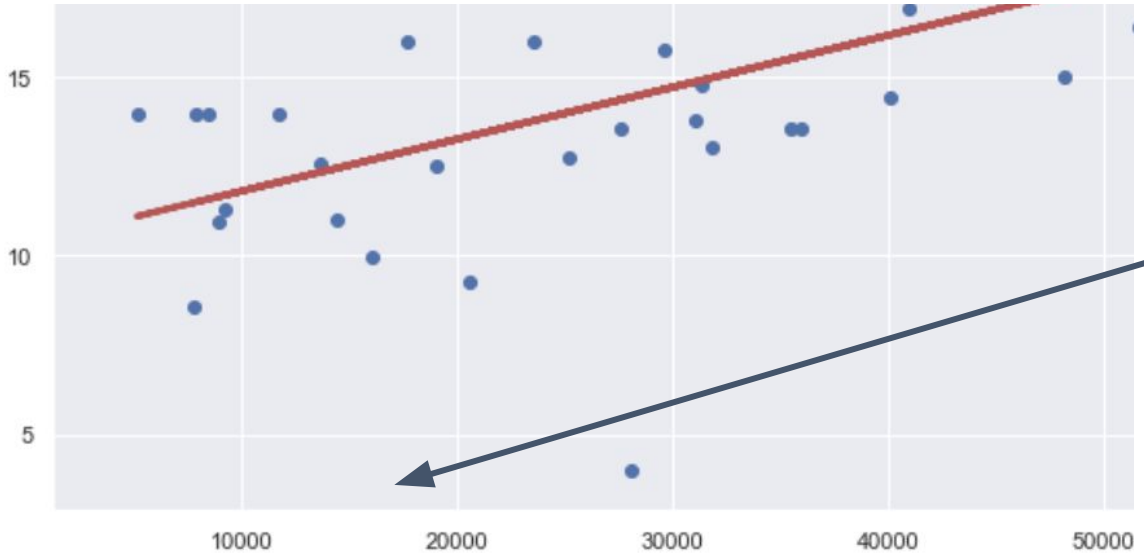
BASIC **STANDARD** **PREMIUM**

HOW DOES A COUNTRY'S GDP AFFECT THE PRICE OF NETFLIX SUBSCRIPTION.?

CORRELATION BETWEEN GDP, PRICES, AND POPULATION DENSITY

NETFLIX

	Basic cost	Standard cost	Premium cost
Intercepts	[6.01838513]	[8.39215105]	[10.35902564]
Coefficients	[[6.12046285e-05]]	[[9.74117271e-05]]	[[0.00014543]]



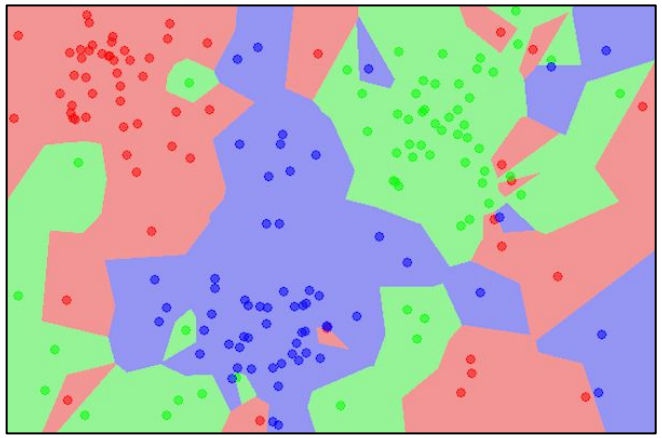
VALUES RANGES BY 10K FOR GDP

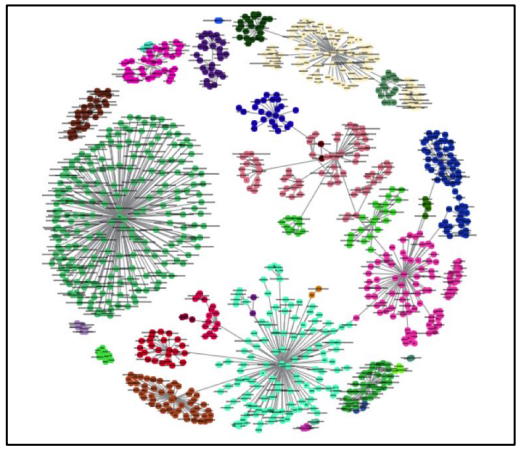
NETFLIX

MACHINE LEARNING

CAN THE PRICE OF A NETFLIX SUBSCRIPTION BE DETERMINED BY JUST GDP AND POPULATION DENSITY? (K-NEAREST NEIGHBOUR)

WHAT IS K-NEAREST NEIGHBOUR





- K-Nearest Neighbour Regression/Classification
- “Bird of a feather flock together”
- Supervised Learning Algorithm
- Easy to learn & Effective

NETFLIX

```
****************** KNN Regression ******************

Test Accuracy Score:  0.14951658687738745
Training Accuracy Score:  0.2629718392933811
----------------------------------------------------------
```

BASIC (K-VALUE: 14)

```
****************** KNN Regression ******************

Test Accuracy Score:  0.2860640866597264
Training Accuracy Score:  0.44508801456093183
----------------------------------------------------------
```

STANDARD (K-VALUE: 13)

```
****************** KNN Regression ******************

Test Accuracy Score:  0.4670006759503321
Training Accuracy Score:  0.608733278943969
----------------------------------------------------------
```

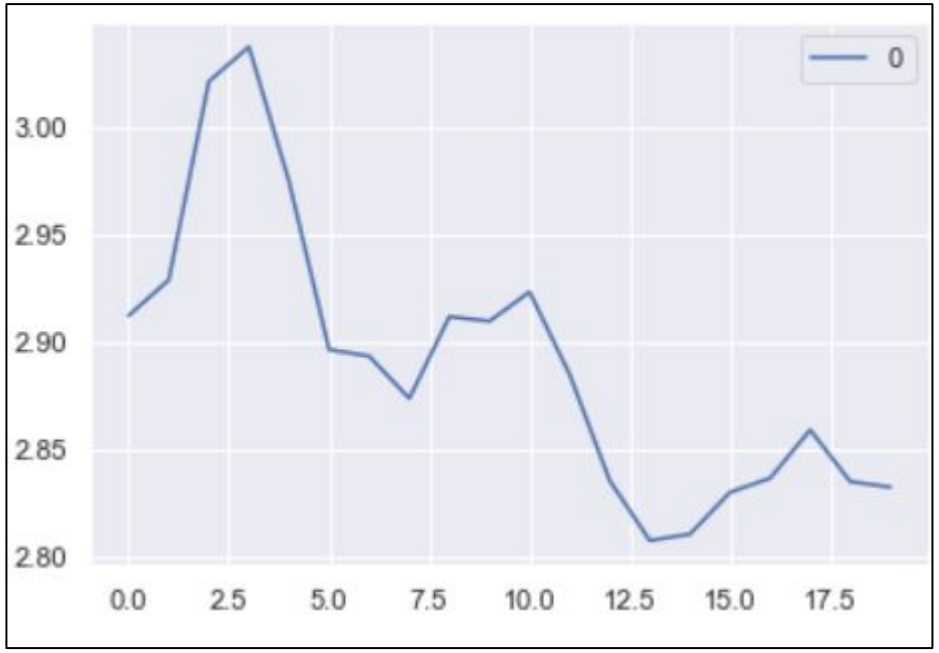
PREMIUM (K-VALUE: 5)

NETFLIX

CHOOSING THE RIGHT K-VALUE

```
rmse_val = []
for K in range(20):
    K = K+1
    model = neighbors.KNeighborsRegressor(n_neighbors = K)

    model.fit(X_train, y_train)
    pred=model.predict(X_test)
    error = sqrt(mean_squared_error(y_test,pred))
    rmse_val.append(error)
    print('RMSE value for k= ' , K , 'is:', error)
```

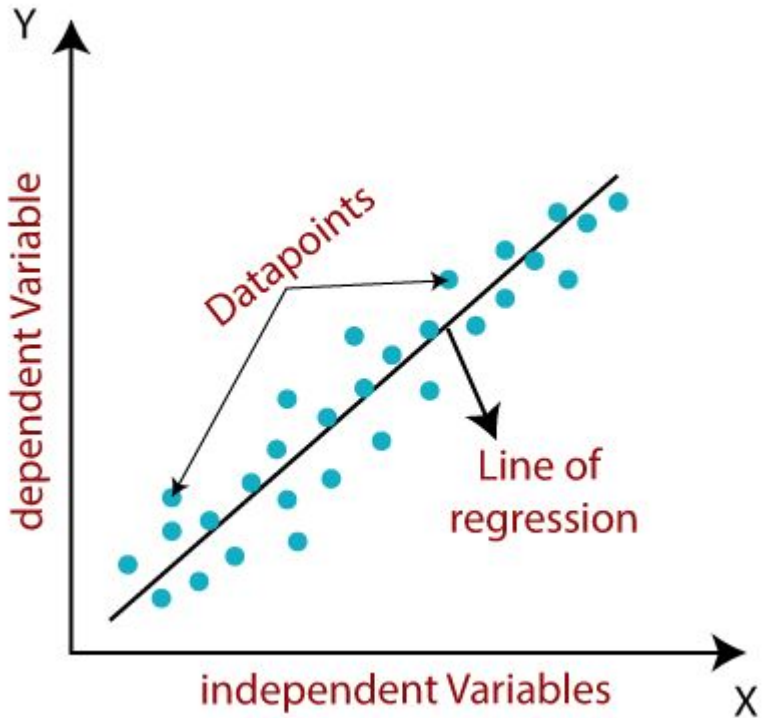
```
RMSE value for k=  1 is: 2.9123587233260486
RMSE value for k=  2 is: 2.9289147592012075
RMSE value for k=  3 is: 3.021373553228423
RMSE value for k=  4 is: 3.0374692213804133
RMSE value for k=  5 is: 2.9747536592688366
RMSE value for k=  6 is: 2.8965041988704376
RMSE value for k=  7 is: 2.8934882027739177
RMSE value for k=  8 is: 2.873779428405214
RMSE value for k=  9 is: 2.911865271976009
RMSE value for k=  10 is: 2.9097361134645867
RMSE value for k=  11 is: 2.9233790215401876
RMSE value for k=  12 is: 2.884981936765456
RMSE value for k=  13 is: 2.8354834109142697
RMSE value for k=  14 is: 2.807624042425453
RMSE value for k=  15 is: 2.8106381969357197
RMSE value for k=  16 is: 2.8300902989587486
RMSE value for k=  17 is: 2.836665363836533
RMSE value for k=  18 is: 2.8591590608043176
RMSE value for k=  19 is: 2.8350967599082884
RMSE value for k=  20 is: 2.832496561486586
```



NETFLIX

CAN THE PRICE OF A NETFLIX SUBSCRIPTION BE DETERMINED BY JUST GDP AND POPULATION DENSITY? (LINEAR REGRESSION/ RIDGE REGRESSION)

LINEAR REGRESSION



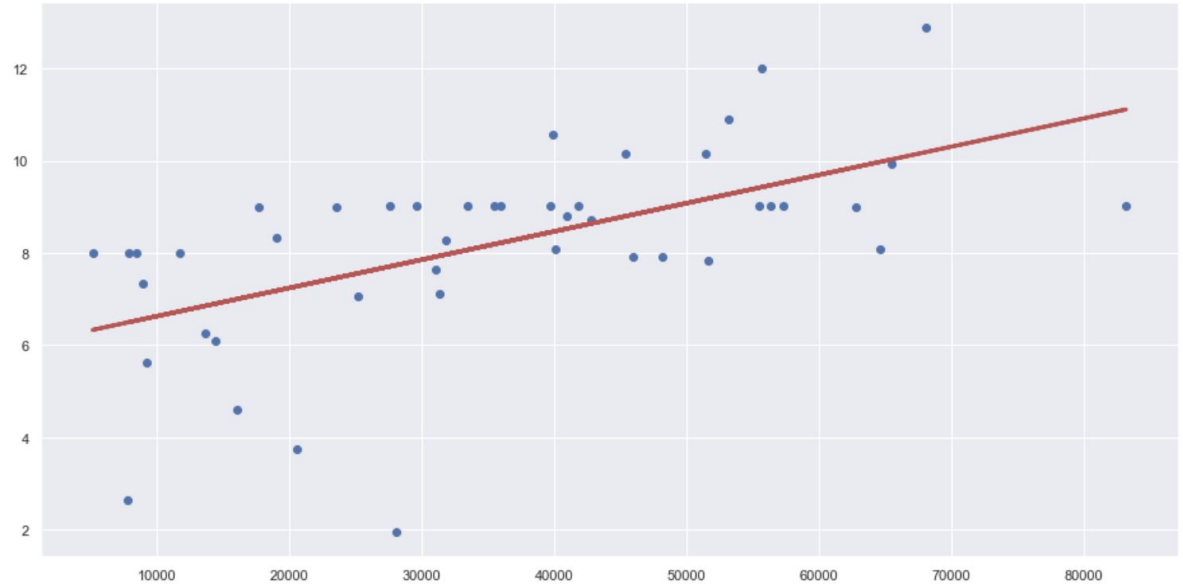
NETFLIX

```python
from sklearn.metrics import accuracy_score

basicPred = linreg1.predict(gdpTrain)
standardPred = linreg2.predict(gdpTrain)
premiumPred = linreg3.predict(gdpTrain)

basictestPred = linreg1.predict(gdpTest)
standardtestPred = linreg2.predict(gdpTest)
premiumtestPred = linreg3.predict(gdpTest)

def mean_sq_err(actual, predicted):
    return np.mean(np.square(np.array(actual) - np.array(predicted)))

mse1 = mean_sq_err(basicTrain, basicPred)
mse2 = mean_sq_err(standardTrain, standardPred)
mse3 = mean_sq_err(premiumTrain, premiumPred)

mse4 = mean_sq_err(basicTest, basictestPred)
mse5 = mean_sq_err(standardTest, standardtestPred)
mse6 = mean_sq_err(premiumTest, premiumtestPred)
print("Mean Squared Error (Train) (MSE) \t:", mse1, mse2, mse3)
print("Mean Squared Error (Test) (MSE) \t:", mse4, mse5, mse6)
print("Root Mean Squared Error (Train) (RMSE) \t:)", np.sqrt(mse1), np.sqrt(mse2), np.sqrt(mse3))
print("Root Mean Squared Error (Test) (RMSE) \t:)", np.sqrt(mse4), np.sqrt(mse5), np.sqrt(mse6))
```

```
Mean Squared Error (Train) (MSE)        : 2.8646851473886388 5.153800044721061 9.127414565487305
Mean Squared Error (Test) (MSE)         : 2.8684904750967246 4.533325265706032 9.766141756919284
Root Mean Squared Error (Train) (RMSE)  :) 1.6925380785638586 2.270198239079808 3.0211611286866686
Root Mean Squared Error (Test) (RMSE)   :) 1.6936618538234616 2.129160695134595 3.125082680013328
```

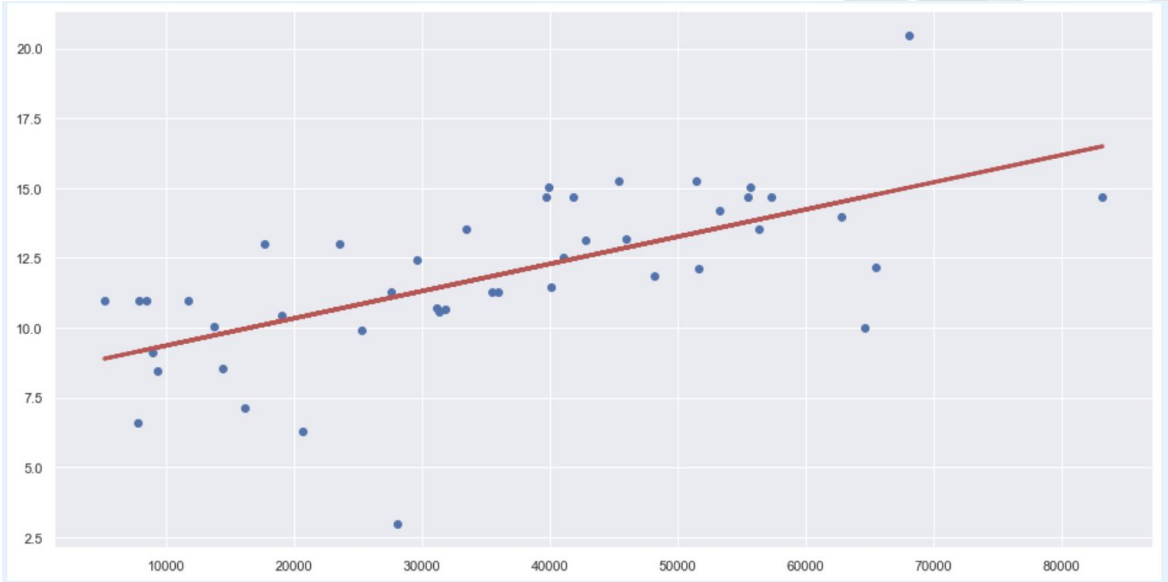
CHECKING FOR OVERFITTING BY COMPARING MSE VALUES BETWEEN TRAIN AND TEST VALUES

NETFLIX



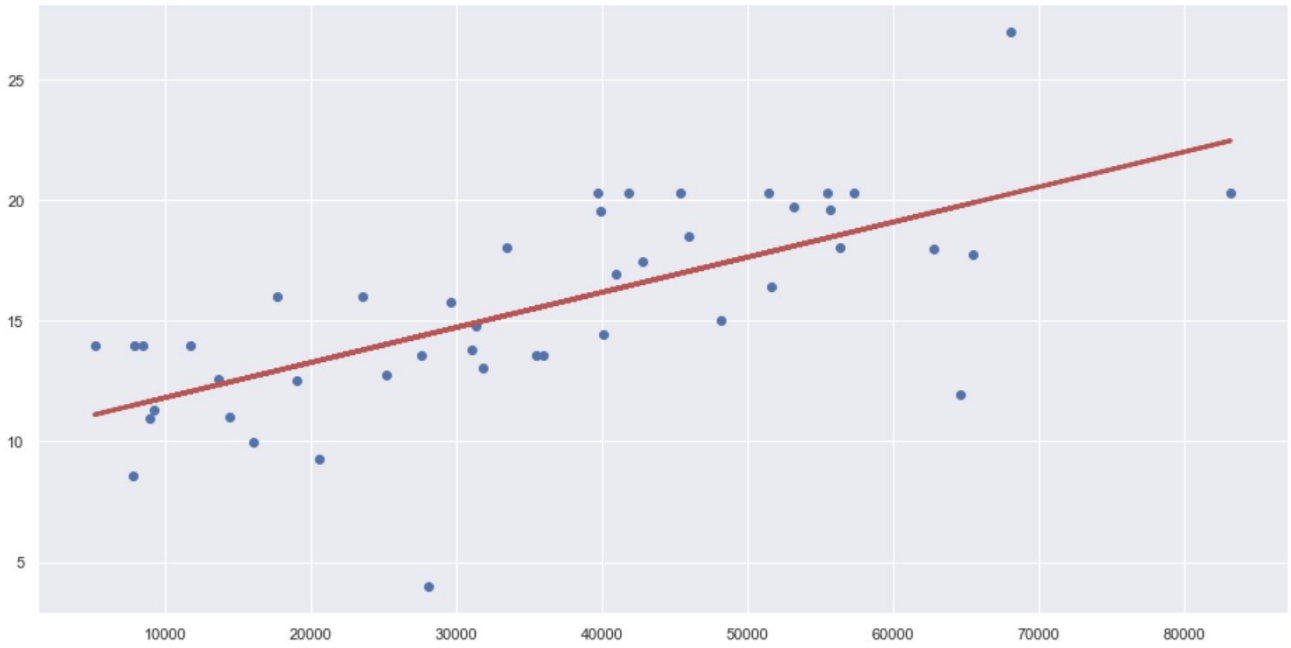
Explained Variance (R^2) : 0.32573255365383613

BASIC



Explained Variance (R^2) : 0.4048303332061981

STANDARD



PREMIUM

NETFLIX

LOOKING AT R^2 VALUES

	BASIC	STANDARD	PREMIUM
R^2 value for GDP	0.32573255365383613	0.4048303332061981	0.46121376160032745
R^2 value for Population density	1.4348364674621195e-05	0.006095914721338302 5	0.013565911841476819

NETFLIX

CONCLUSION

PRICES OF NETFLIX INTERNATIONALLY

- GDP is an important factor in pricing of Netflix or even goods internationally
- Purchasing Power parity is not omnipotent as there are many other factors in play
- Taxes, imports, competition, government, laws are all factors that need to be considered as well
- Thus, we are unable to solve the question of "How are the prices of various Netflix subscriptions determined in a country?" as there are too many other considerations unknown to public

THANK YOU!