EDS Mini Project -(G Division)

Group members name: Kanishka Garud – 723

Apurva Koli – 736

Shravani Halkude - 727

Problem Statement:

Implement a mini project based onclassification (Linear Regression / KNN Classification) or Clustering (K-Means) and also Develop an interactive
dashboard using the matplotlib/Seaborn library.

Data set:

A	B	C	D	E
Brand	Category	Price	Color	Size
Zara	T-shirt	19.99	Black	S
H&M	Jeans	39.99	Blue	M
GAP	Hoodie	29.99	Gray	L
Forever 2:	Dress	24.99	Red	S
Nike	Shoes	79.99	White	8
Adidas	T-shirt	29.99	Blue	L
Levi's	Jeans	49.99	Black	32
Puma	Shorts	19.99	Gray	XL
Calvin Kle	Underwea	14.99	Black	M
Tommy Hi	Shirt	34.99	White	M

Code:

```
import numpy as nm
i
```

	Brand	Category	Price	Color	Size
0	Zara	T-shirt	19.99	Black	S
1	H&M	Jeans	39.99	Blue	M
2	GAP	Hoodie	29.99	Gray	L
3	Forever 21	Dress	24.99	Red	S
4	Nike	Shoes	79.99	White	8
5	Adidas	T-shirt	29.99	Blue	L
6	Levi's	Jeans	49.99	Black	32
7	Puma	Shorts	19.99	Gray	XL
8	Calvin Klein	Underwear	14.99	Black	M
9	Tommy Hilfiger	Shirt	34.99	White	M

```
x
=
d
```

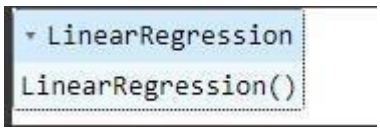
```
[['H&M' 'Jeans' 39.99 'Blue']
 ["Levi's" 'Jeans' 49.99 'Black']
 ['Puma' 'Shorts' 19.99 'Gray']
 ['Forever 21' 'Dress' 24.99 'Red']
 ['Zara' 'T-shirt' 19.99 'Black']
 ['Adidas' 'T-shirt' 29.99 'Blue']]
```

Code: Linear Regression

```
#
```

```
data_set

df = pd.DataFrame(data_set)
```

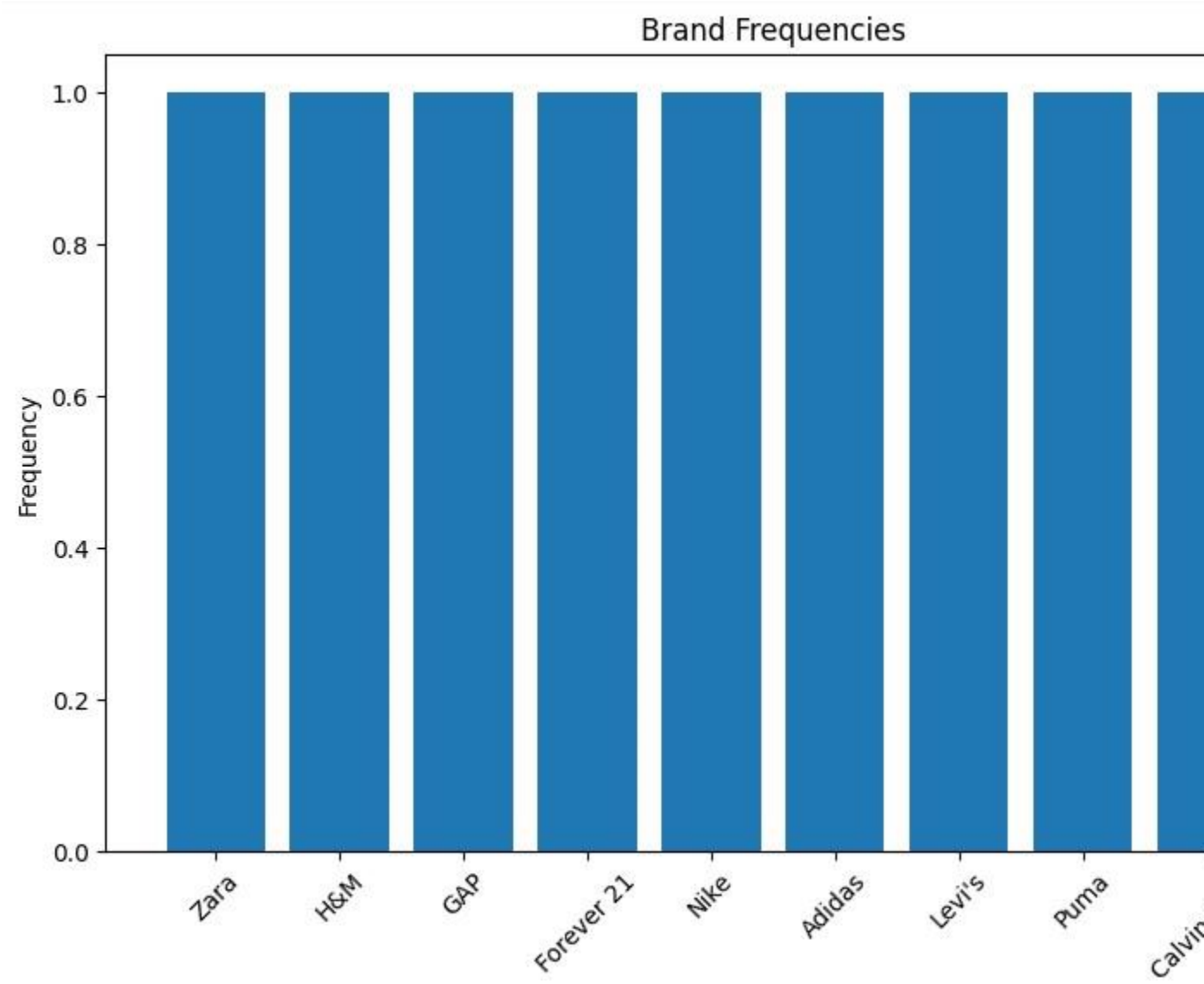



```
#print the coefficient
```

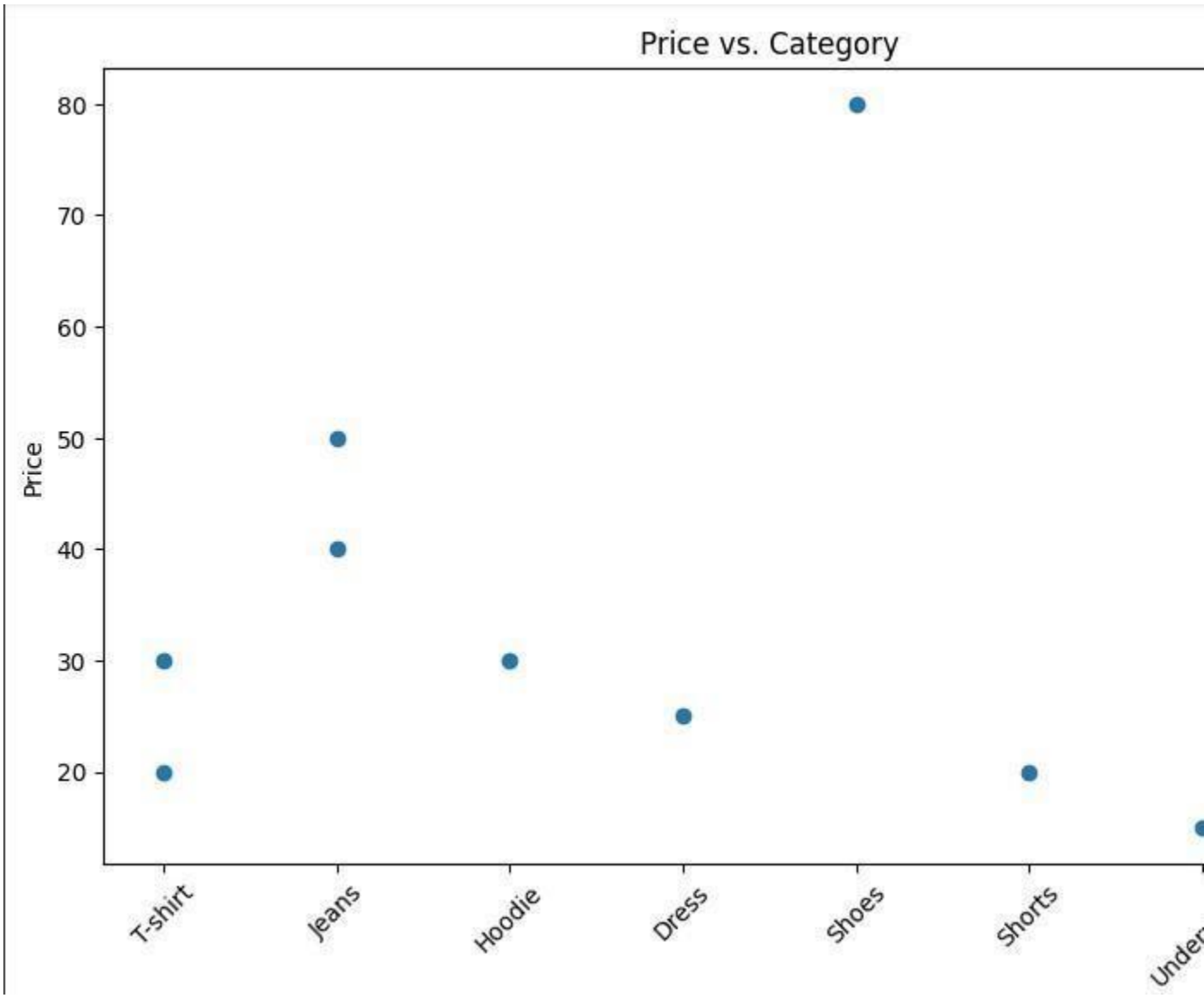
```
Intercept: 7.105427357601002e-15
Coefficients: [ 8.09593144e-16  1.66533454e-15  1.00000000e+00  7.35522754e-16
 -2.59514632e-15]
```

Code:Visualization

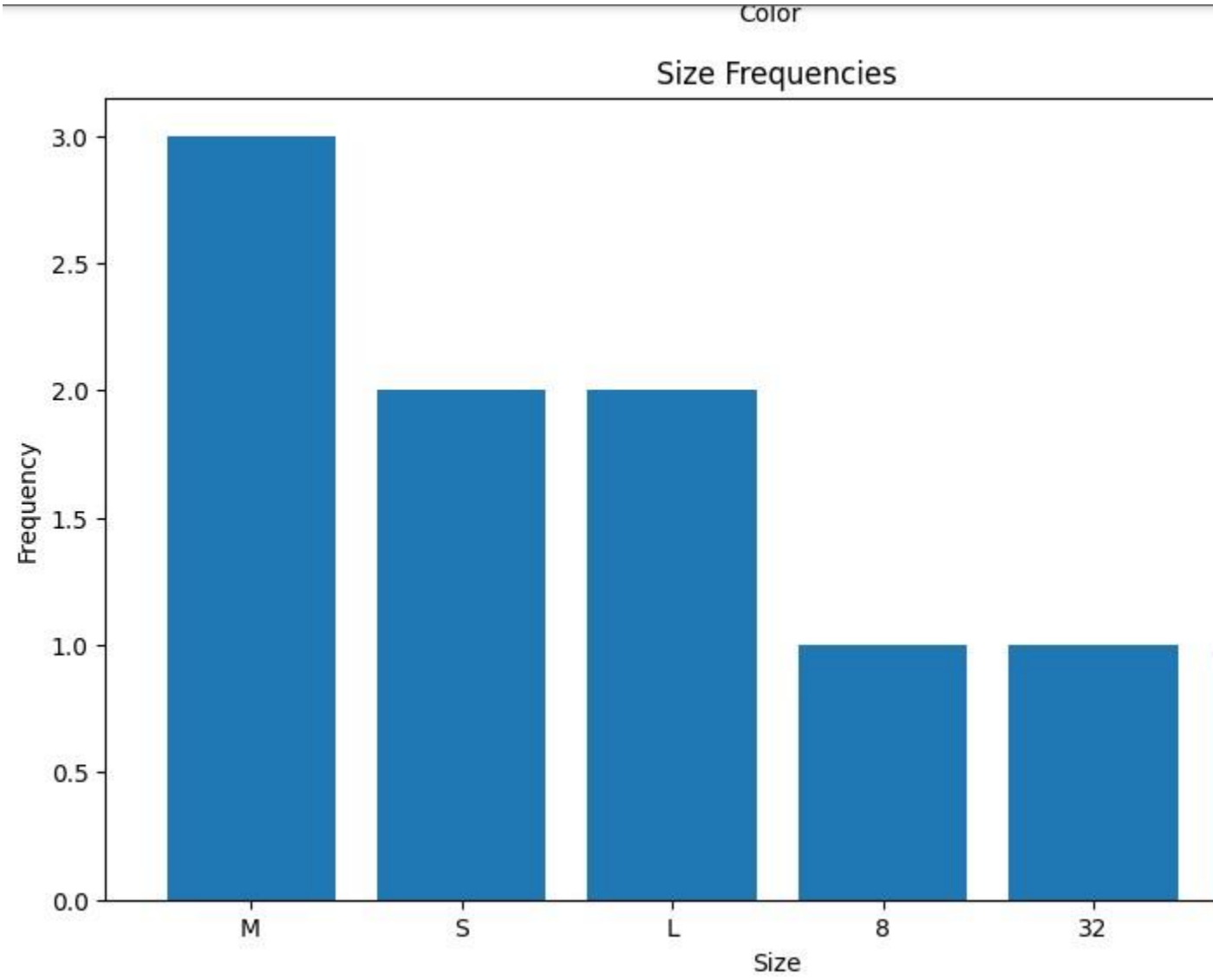
```
#visaulization
#
b
a
```




```
# Scatter plot of price vs. category
plt.figure(figsize=(10, 6))
plt.scatter(df['Category'], df['Price'])
plt.xlabel('Category')
plt.ylabel('Price')
plt.title('Price vs. Category')
plt.xticks(rotation=45)
plt.show()
```




```
# Bar plot of size frequencies
size_counts = df['Size'].value_counts()
plt.figure(figsize=(10, 6))
plt.bar(size_counts.index, size_counts.values)
plt.xlabel('Size')
plt.ylabel('Frequency')
plt.title('Size Frequencies')
plt.show()
```



Code: Manupulation

```
#
m
a
n
u
```

	Brand	Price
0	Zara	19.99
1	H&M	39.99
2	GAP	29.99
3	Forever 21	24.99
4	Nike	79.99
5	Adidas	29.99
6	Levi's	49.99
7	Puma	19.99
8	Calvin Klein	14.99
9	Tommy Hilfiger	34.99

```
#
```

	Brand	Category	Price	Color	Size
1	H&M	Jeans	39.99	Blue	M
4	Nike	Shoes	79.99	White	8
6	Levi's	Jeans	49.99	Black	32
9	Tommy Hilfiger	Shirt	34.99	White	M

```
# Sort the dataframe by a
```

	Brand	Category	Price	Color	Size
4	Nike	Shoes	79.99	White	8
6	Levi's	Jeans	49.99	Black	32
1	H&M	Jeans	39.99	Blue	M
9	Tommy Hilfiger	Shirt	34.99	White	M
2	GAP	Hoodie	29.99	Gray	L
5	Adidas	T-shirt	29.99	Blue	L
3	Forever 21	Dress	24.99	Red	S
0	Zara	T-shirt	19.99	Black	S
7	Puma	Shorts	19.99	Gray	XL
8	Calvin Klein	Underwear	14.99	Black	M

```
#
```

Category	Price
Dress	24.99
Hoodie	29.99
Jeans	44.99
Shirt	34.99
Shoes	79.99
Shorts	19.99
T-shirt	24.99
Underwear	14.99

```
#
```

	Brand	Category	Price	Color	Size
0	Zara	T-shirt	19.99	Black	S
1	H&M	Jeans	39.99	Blue	M
2	GAP	Hoodie	29.99	Gray	L
3	Forever 21	Dress	24.99	Red	S
4	Nike	Shoes	79.99	White	8
5	Adidas	T-shirt	29.99	Blue	L
6	Levi's	Jeans	49.99	Black	32
7	Puma	Shorts	19.99	Gray	XL
8	Calvin Klein	Underwear	14.99	Black	M
9	Tommy Hilfiger	Shirt	34.99	White	M

Code: K-means clustering

```
#k-means clustering
#

P
r
e
p
r
```

```
#
Add
the
clus
ter
```

```
   Brand  Cluster
0      9        1
1      4        0
2      3        1
3      2        1
4      6        2
5      0        1
6      5        0
7      7        1
8      1        1
9      8        0
Cluster Centers:
Cluster 1 Center: [41.65666667  1.66666667  2.        ]
Cluster 2 Center: [23.32333333  1.33333333  3.33333333]
Cluster 3 Center: [79.99  4.    1.  ]
```

Code: KNN clasification

```
#KNN clasification
#

P
r
```

```
X_train, X_test,
y_train, y_test =
train_test_split(X,
```

```
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=3)
```

```
#
```

```
Accuracy: 0.0
```