In [1]:

```python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

In [3]:

```python
# Uploading Csv file
df = pd.read_csv(r"Billionaires.csv")

# Data Preprocessing - .head()
df.head()
```

Out[3]:

	rank	finalWorth	category	personName	age	country	city	source
0	1	211000	Fashion & Retail	Bernard Arnault & family	74.0	France	Paris	LVMH
1	2	180000	Automotive	Elon Musk	51.0	United States	Austin	Tesla, SpaceX
2	3	114000	Technology	Jeff Bezos	59.0	United States	Medina	Amazon
3	4	107000	Technology	Larry Ellison	78.0	United States	Lanai	Oracle
4	5	106000	Finance & Investments	Warren Buffett	92.0	United States	Omaha	Berkshire Hathaway

5 rows × 35 columns

In [4]:

```python
df.tail()
```

Out[4]:

	rank	finalWorth	category	personName	age	country	city	
2635	2540	1000	Healthcare	Yu Rong	51.0	China	Shanghai	I
2636	2540	1000	Food & Beverage	Richard Yuengling, Jr.	80.0	United States	Pottsville	
2637	2540	1000	Manufacturing	Zhang Gongyun	60.0	China	Gaomi	m
2638	2540	1000	Real Estate	Zhang Guiping & family	71.0	China	Nanjing	
2639	2540	1000	Diversified	Inigo Zobel	66.0	Philippines	Makati	

5 rows × 35 columns

In [5]:

```python
df.shape
```

Out[5]:

```
(2640, 35)
```

In [6]:
```
df.columns
```

Out[6]:
```
Index(['rank', 'finalWorth', 'category', 'personName', 'age', 'country',
       'city', 'source', 'industries', 'countryOfCitizenship', 'organizatio
n',
       'selfMade', 'status', 'gender', 'birthDate', 'lastName', 'firstName',
       'title', 'date', 'state', 'residenceStateRegion', 'birthYear',
       'birthMonth', 'birthDay', 'cpi_country', 'cpi_change_country',
       'gdp_country', 'gross_tertiary_education_enrollment',
       'gross_primary_education_enrollment_country', 'life_expectancy_countr
y',
       'tax_revenue_country_country', 'total_tax_rate_country',
       'population_country', 'latitude_country', 'longitude_country'],
      dtype='object')
```

In [7]:
```
df.dtypes
```

Out[7]:
```
rank                                            int64
finalWorth                                      int64
category                                       object
personName                                     object
age                                           float64
country                                        object
city                                           object
source                                         object
industries                                     object
countryOfCitizenship                           object
organization                                   object
selfMade                                         bool
status                                         object
gender                                         object
birthDate                                      object
lastName                                       object
firstName                                      object
title                                          object
date                                           object
state                                          object
residenceStateRegion                           object
birthYear                                     float64
birthMonth                                    float64
birthDay                                      float64
cpi_country                                   float64
cpi_change_country                            float64
gdp_country                                    object
gross_tertiary_education_enrollment           float64
gross_primary_education_enrollment_country    float64
life_expectancy_country                       float64
tax_revenue_country_country                   float64
total_tax_rate_country                        float64
population_country                            float64
latitude_country                              float64
longitude_country                             float64
dtype: object
```

In [8]:
```python
df["category"].unique()
```

Out[8]:
```
array(['Fashion & Retail', 'Automotive', 'Technology',
       'Finance & Investments', 'Media & Entertainment', 'Telecom',
       'Diversified', 'Food & Beverage', 'Logistics',
       'Gambling & Casinos', 'Manufacturing', 'Real Estate',
       'Metals & Mining', 'Energy', 'Healthcare', 'Service',
       'Construction & Engineering', 'Sports'], dtype=object)
```

In [9]:
```python
df.nunique()
```

Out[9]:
```
rank                                          219
finalWorth                                    219
category                                       18
personName                                   2638
age                                            79
country                                        78
city                                          741
source                                        906
industries                                     18
countryOfCitizenship                           77
organization                                  294
selfMade                                        2
status                                          6
gender                                          2
birthDate                                    2060
lastName                                     1736
firstName                                    1770
title                                          97
date                                            2
state                                          45
residenceStateRegion                            5
birthYear                                      77
birthMonth                                     12
birthDay                                       31
cpi_country                                    63
cpi_change_country                             44
gdp_country                                    68
gross_tertiary_education_enrollment            63
gross_primary_education_enrollment_country     60
life_expectancy_country                        54
tax_revenue_country_country                    57
total_tax_rate_country                         63
population_country                             68
latitude_country                               68
longitude_country                              68
dtype: int64
```

In [10]:
```python
df.describe()
```

Out[10]:

	rank	finalWorth	age	birthYear	birthMonth	bi
count	2640.000000	2640.000000	2575.000000	2564.000000	2564.000000	2564.
mean	1289.159091	4623.787879	65.140194	1957.183307	5.740250	12.
std	739.693726	9834.240939	13.258098	13.282516	3.710085	9.
min	1.000000	1000.000000	18.000000	1921.000000	1.000000	1.
25%	659.000000	1500.000000	56.000000	1948.000000	2.000000	1.
50%	1312.000000	2300.000000	65.000000	1957.000000	6.000000	11.
75%	1905.000000	4200.000000	75.000000	1966.000000	9.000000	21.
max	2540.000000	211000.000000	101.000000	2004.000000	12.000000	31.

In [11]:

```
df["category"].value_counts()
```

Out[11]:

```
Finance & Investments         372
Manufacturing                 324
Technology                    314
Fashion & Retail              266
Food & Beverage               212
Healthcare                    201
Real Estate                   193
Diversified                   187
Energy                        100
Media & Entertainment          91
Metals & Mining                74
Automotive                     73
Service                        53
Construction & Engineering     45
Logistics                      40
Sports                         39
Telecom                        31
Gambling & Casinos             25
Name: category, dtype: int64
```

In [12]:

```
df.isnull()
```

Out[12]:

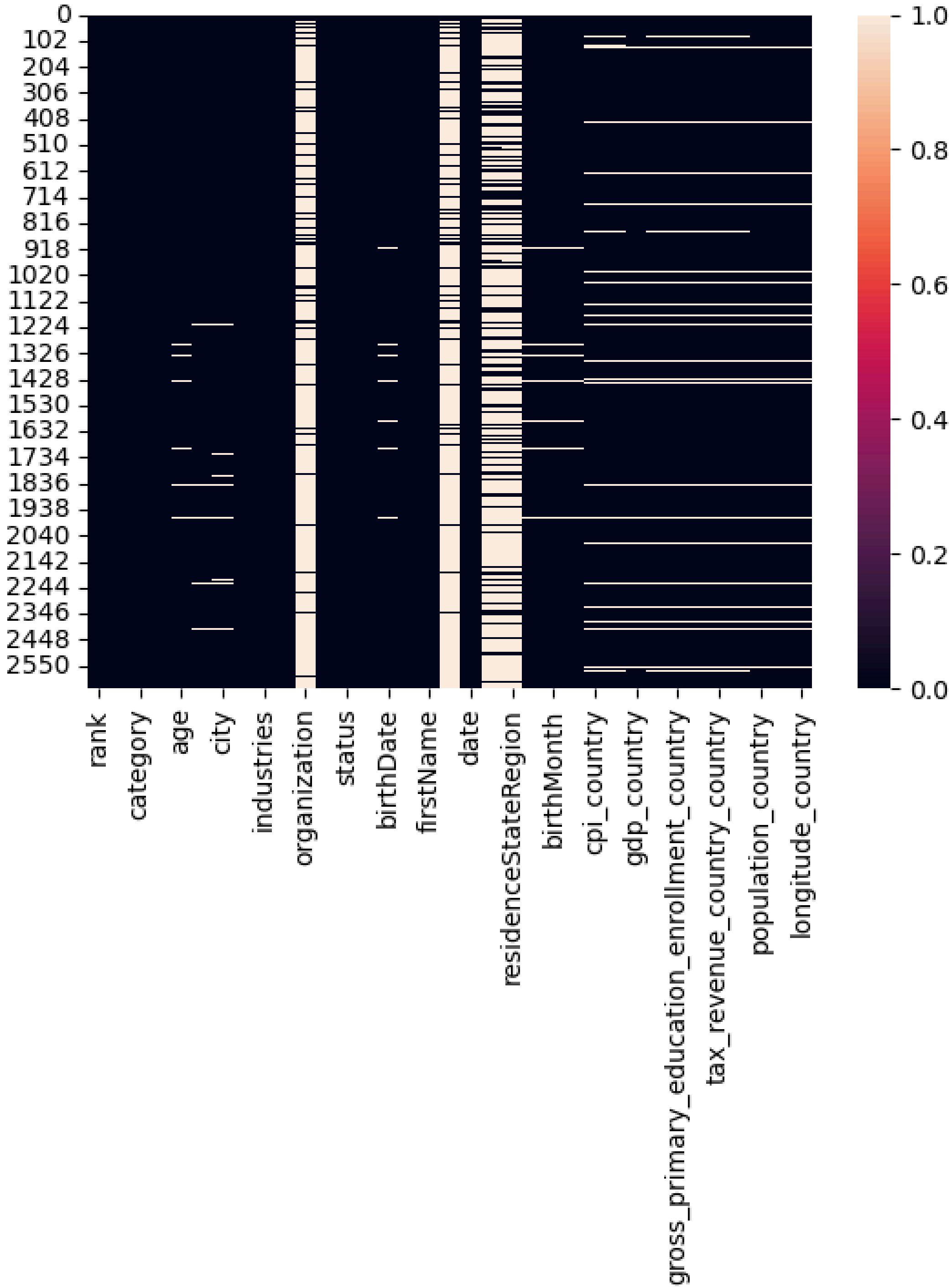
	rank	finalWorth	category	personName	age	country	city	source	in
0	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	
2635	False	False	False	False	False	False	False	False	
2636	False	False	False	False	False	False	False	False	
2637	False	False	False	False	False	False	False	False	
2638	False	False	False	False	False	False	False	False	
2639	False	False	False	False	False	False	False	False	

2640 rows × 35 columns

In [13]:

```
sns.heatmap(df.isnull())
```

Out[13]: <Axes: >



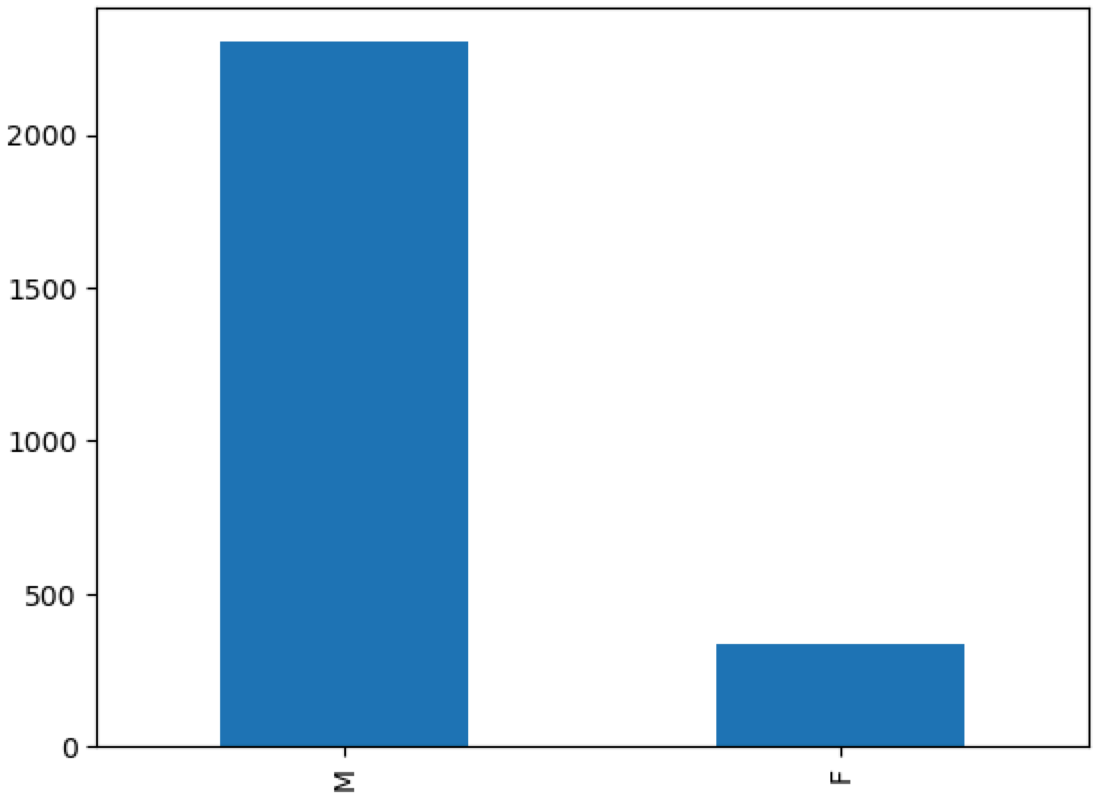
📊 Data Visualization and Analysis

In [14]:
```
# Count of Gender
gender_diversity = df['gender'].value_counts()
gender_diversity
```

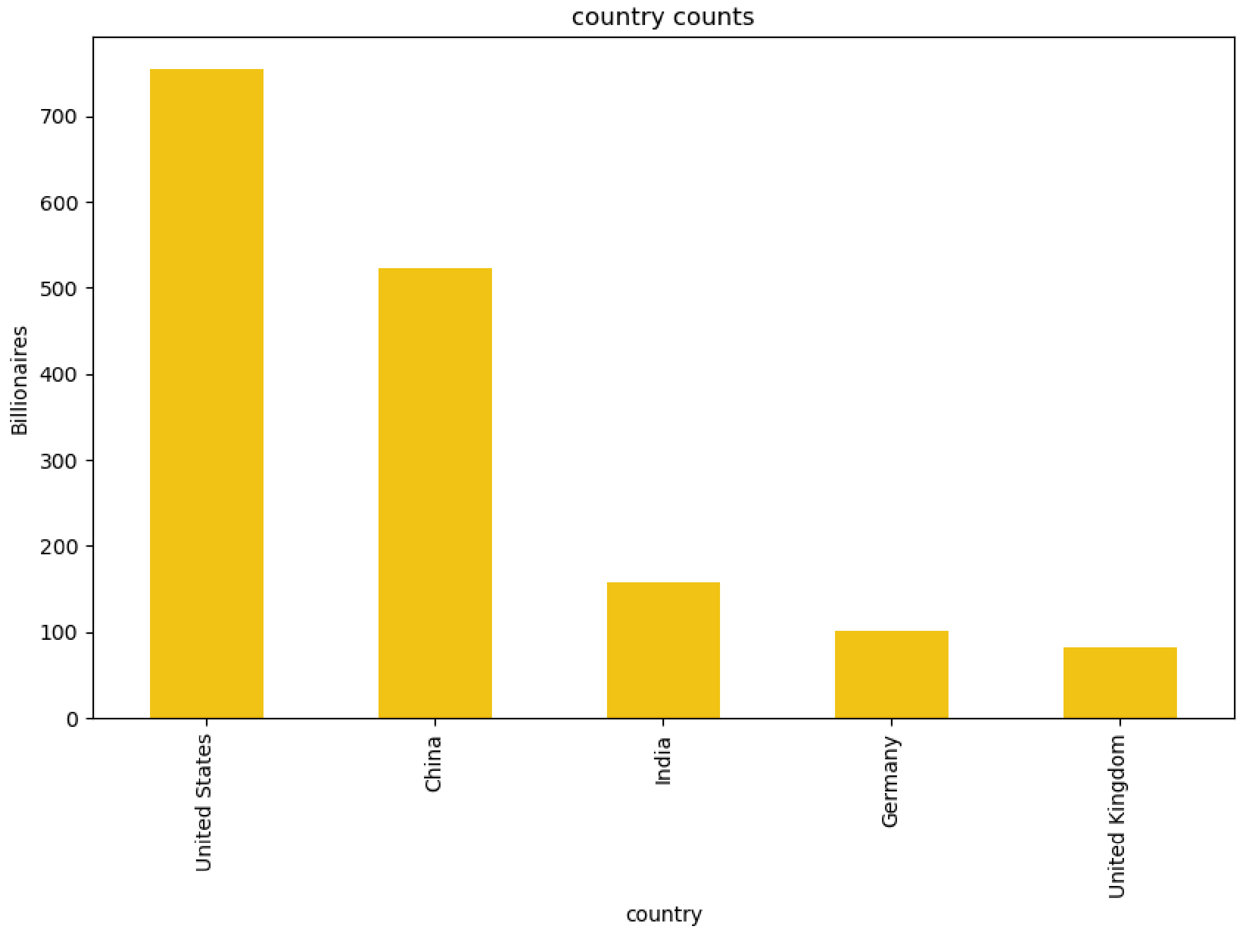
Out[14]:
```
M    2303
F     337
Name: gender, dtype: int64
```

In [16]:
```
# Plotting Gender Count
df.gender.value_counts().plot(kind= 'bar')
```

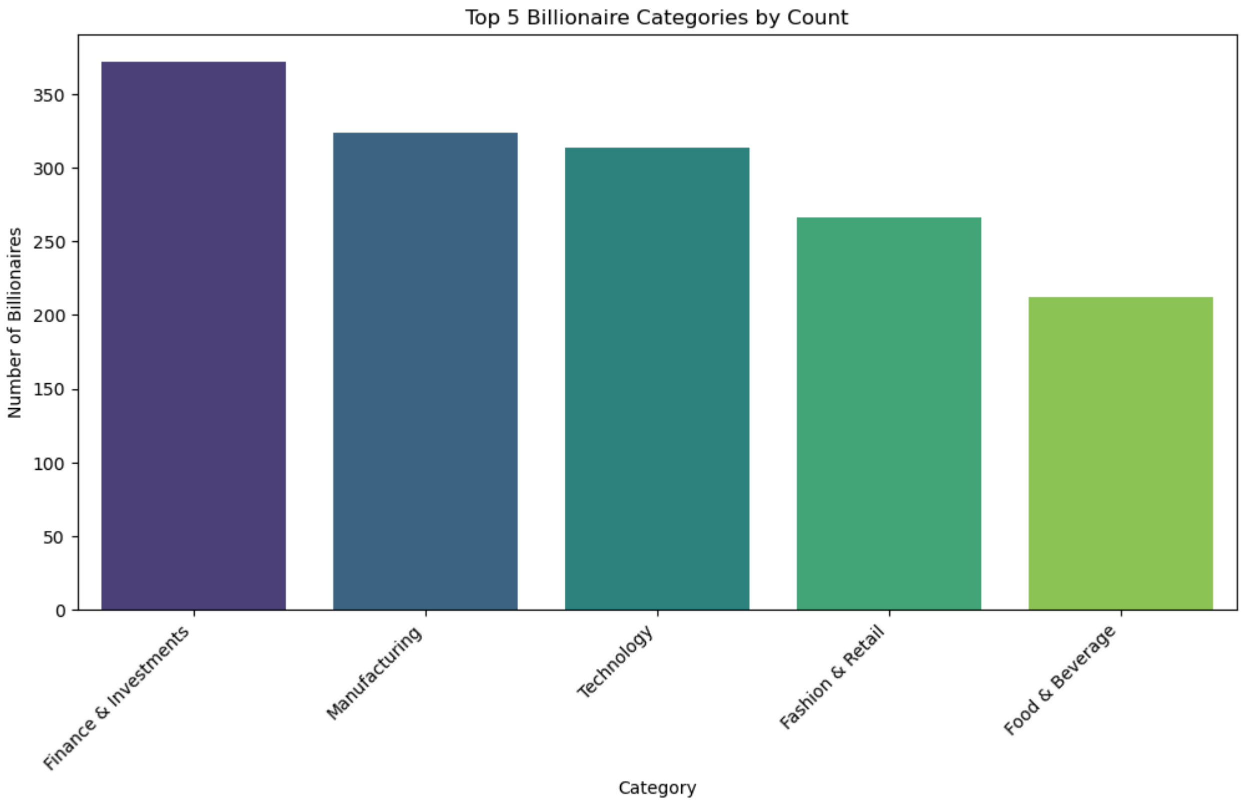
Out[16]: <Axes: >



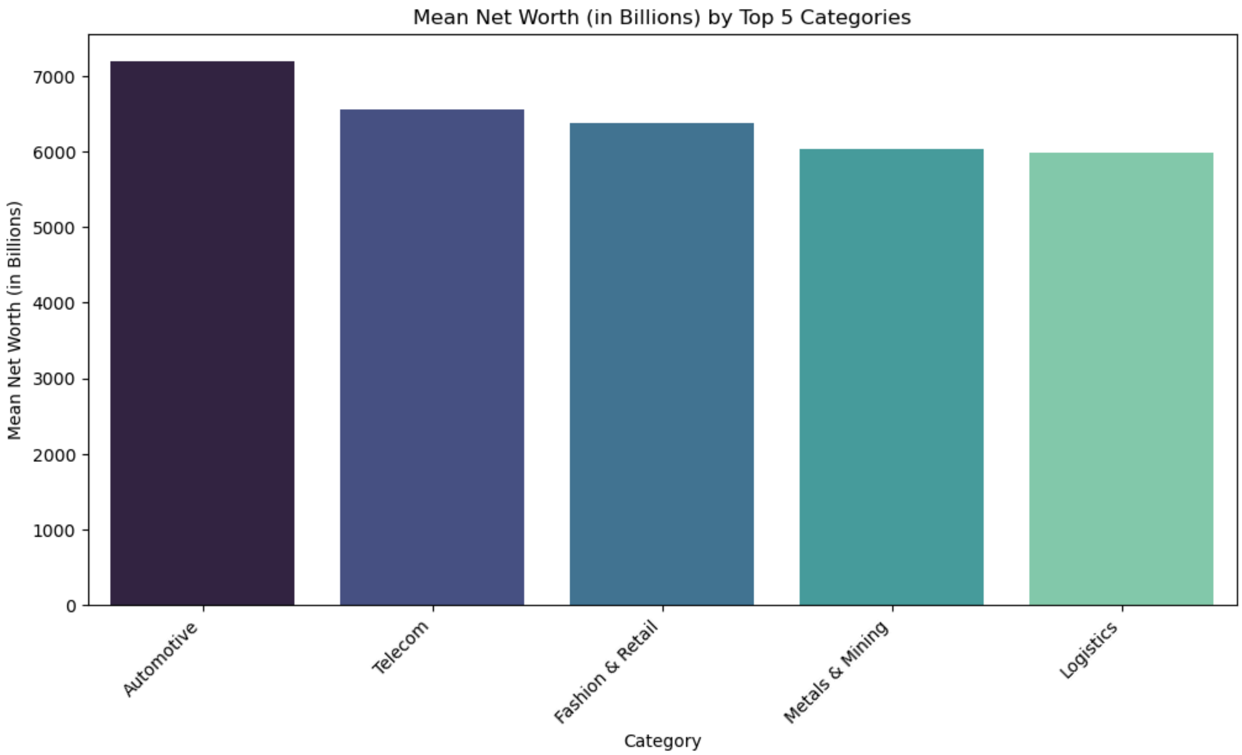
In [17]:
```
# Country Wise Billionaires (Top 5)
country_counts = df["country"].value_counts().head()
plt.figure(figsize=(10,6))
country_counts.plot(kind='bar', color = "#F4C314")
plt.title('country counts')
plt.xlabel('country')
plt.ylabel('Billionaires')
plt.xticks(rotation = 90)
plt.show()
```



In [18]:
```
category_counts = df["category"].value_counts().head(5)

plt.figure(figsize=(12, 6))
# Create a bar plot for the top 5 categories
sns.barplot(x=category_counts.index, y=category_counts.values, palette='viridi
plt.title('Top 5 Billionaire Categories by Count')
plt.xlabel('Category')
plt.ylabel('Number of Billionaires')
plt.xticks(rotation=45, ha='right')
plt.show()
```

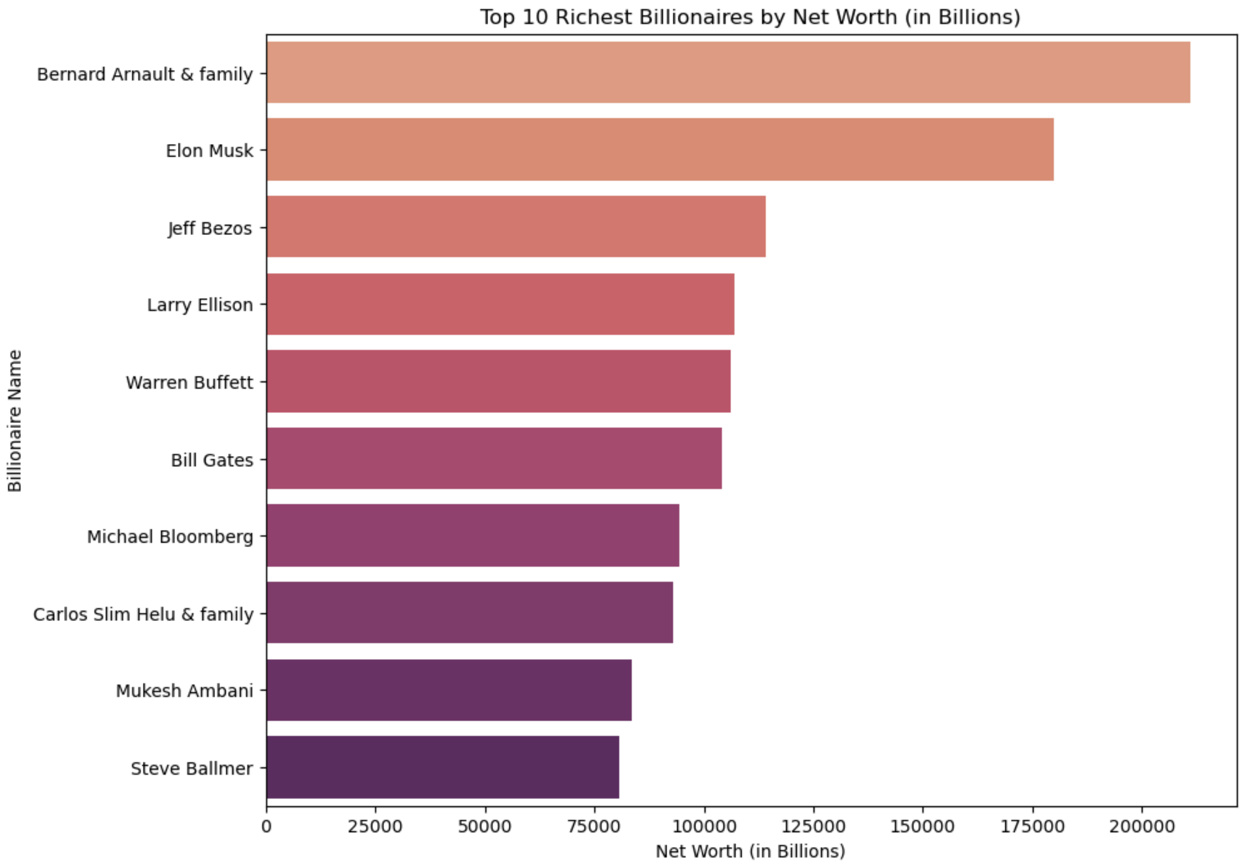


In [20]:
```
# Calculating the mean net worth for each category
mean_wealth_by_category = df.groupby('category')['finalWorth'].mean().sort_val

plt.figure(figsize=(12, 6))
# Create a bar plot
sns.barplot(x=mean_wealth_by_category.index, y=mean_wealth_by_category.values,
plt.title('Mean Net Worth (in Billions) by Top 5 Categories')
plt.xlabel('Category')
plt.ylabel('Mean Net Worth (in Billions)')
plt.xticks(rotation=45, ha='right')
plt.show()
```



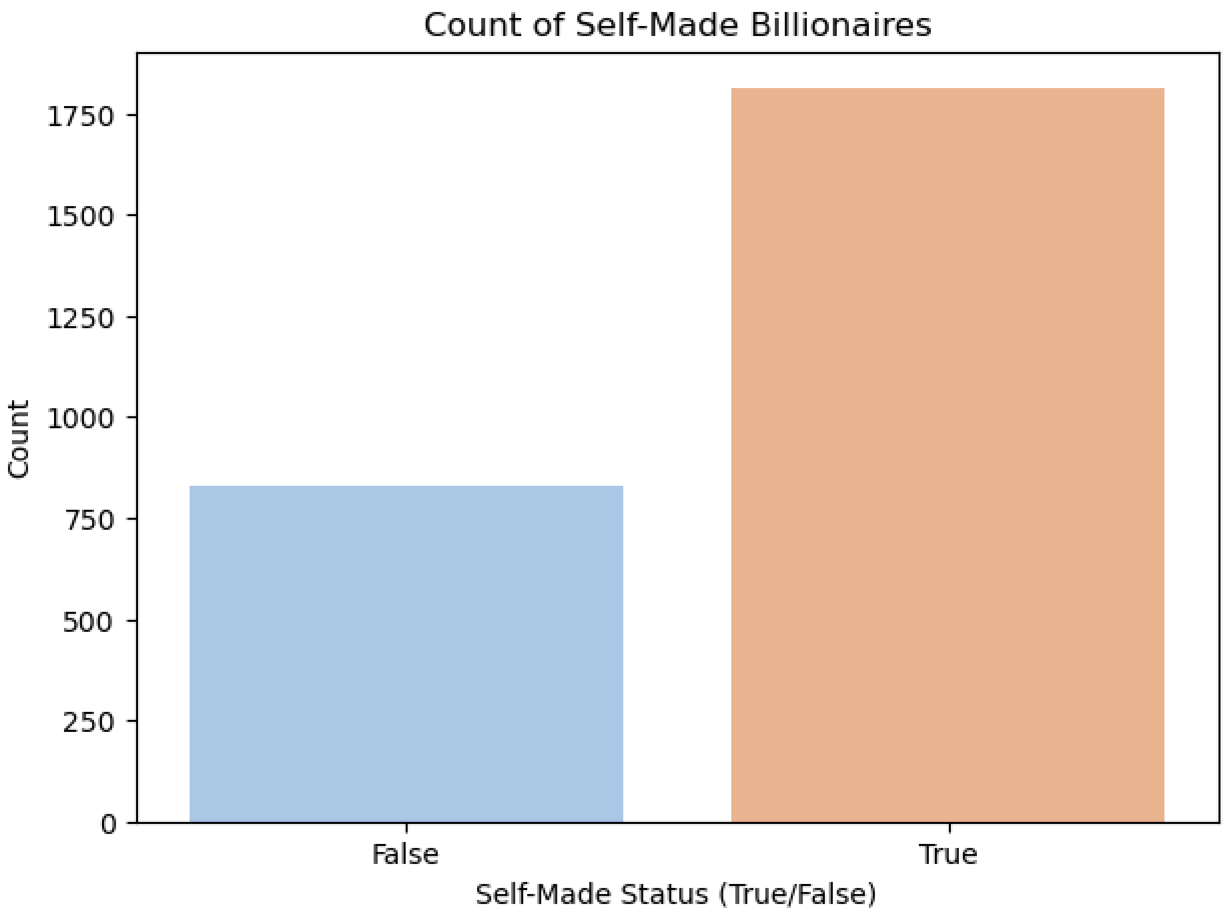
In [22]:

```
 # Sorting the DataFrame by 'fina; Worth' in descending order and select the t
top_10_billionaires = df.sort_values(by='finalWorth', ascending=False).head(10

plt.figure(figsize=(10, 8))
# Creating a horizontal bar plot for better readability of names
sns.barplot(x='finalWorth', y='personName', data=top_10_billionaires, palette=
plt.title('Top 10 Richest Billionaires by Net Worth (in Billions)')
plt.xlabel('Net Worth (in Billions)')
plt.ylabel('Billionaire Name')
plt.show()
```



In [23]:

```
# Plotting the count of Self-Made vs. Not Self-Made billionaires
plt.figure(figsize=(7, 5))
sns.countplot(x='selfMade', data=df, palette='pastel')
plt.title('Count of Self-Made Billionaires')
plt.xlabel('Self-Made Status (True/False)')
plt.ylabel('Count')
plt.show()
```


Count of Self-Made Billionaires
1750
1500
1250
1000
750
500
250
0
Count
False
True
Self-Made Status (True/False)

In [ ]:

Main findings (data + visuals)

Geographic concentration — US and China dominate.
The country bar chart shows the United States (~700 entries) and China (~500 entries) far ahead of other countries (India, Germany, UK are much smaller). This indicates global billionaire counts are concentrated in a few countries.

Industry concentration — finance, manufacturing and tech are largest by headcount.
The category counts list top counts as Finance & Investments, Manufacturing, Technology, Fashion & Retail, and Food & Beverage — these five categories account for a large share of the dataset. This suggests these sectors are the main sources of billionaire creation.

Strong gender imbalance.
Gender counts show M: 2303 vs F: 337 — women make up only ~12–13% of the billionaires in this dataset, showing a clear gender gap.

Majority are self-made.
selfMade counts (page 15) show True: 1812, False: 828 — about 68.6% self-made, 31.4% inherited. This indicates most billionaires in the dataset created their own wealth, though a sizeable inherited share remains.

Net wealth is heavily right-skewed — a few individuals hold outsized wealth.
The describe() output and the “Top billionaires” table show the largest values (e.g., Bernard Arnault 211000; Elon Musk 180000; Jeff Bezos 114000; …). The top 10 listed sum to 1,174,400 (dataset units), illustrating heavy concentration at the very top and a mean net worth driven upward by a small group. The visual bar of the Top-10confirms this concentration.

Mean net worth differs by industry.
The “Mean Wealth by Category” bar chart shows clear differences: some industries (notably finance/investment and several tech/retail categories) have higher average net worth per billionaire than others (e.g., food/beverage appears lower). This suggests industry matters not just for headcount but for average wealth size.