

Top Intelligent People in the World

In this project, I conduct an exploratory data analysis on a dataset featuring some of the most intelligent individuals around the globe. The analysis explores various attributes such as IQ scores, notable achievements, professions, and contributions to society. By visualizing and interpreting this data, we aim to uncover patterns, trends, and insights that define these extraordinary minds. This project provides a deeper understanding of the correlation between intelligence, impact, and recognition on a global scale.

```
In [121]: from IPython.display import Image
display(Image(filename="r'C:\Users\Syed Arif\OneDrive\Desktop\TopIntelligentPeople.jpg"))

Out[121]:
```

Import Library

```
In [5]: import pandas as pd

In [7]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

Uploading Csv file

```
In [9]: df = pd.read_csv("r'C:\Users\Syed Arif\OneDrive\Desktop\TopIntelligentPeople_in_the_world_5000.csv")
```

Data Preprocessing

head()

head is used show to the By default = 5 rows in the dataset

```
In [15]: df.head()
```

	Name	Country	Field of Expertise	IQ	Achievements	Birth Year	Gender	Notable Works	Awards	Education	Influence
0	Enrico Fermi	Austria	Polymath	199	Father of Computer Science	1968	Female	E=mc²	Numerous Posthumous	Self-taught	Popularizing science and cosmology
1	Hilax Plarck	Italy	Chemistry	159	Theory of Evolution	1980	Female	Bohr Model	Nobel Prize	Ph.D. in Astronomy	Foundational work in quantum mechanics
2	Paul Dirac	UK	Physics	177	Quantum Mechanics	1927	Female	Cosmos	Nobel Prize	Ph.D. in Mathematics	Foundation of classical mechanics
3	Erwin Schrodinger	Italy	Physics	130	Electromagnetic Induction	1921	Female	Discovery of Electromagnetic Induction	Nobel Prize	University of Cambridge	Isonic Renaissance artist and inventor
4	Paul Dirac	UK	Physics	163	Wave Equation	1964	Female	On Computable Numbers	Nobel Prize	Ph.D. in (honorary)	Foundational work in quantum mechanics

.tail()

tail is used to show rows by Descending order

```
In [17]: df.tail()
```

```
Out[17]:
```

	Name	Country	Field of Expertise	IQ	Achievements	Birth Year	Gender	Notable Works	Awards	Education	Influence
4995	Michael Faraday	UK	Mathematics	166	Theory of Relativity	1957	Female	Various Patents	NaN	University of Pisa	Foundational work in quantum mechanics
4996	James Clerk Maxwell	USA	Physics	162	Electromagnetic Induction	1926	Female	Bohr Model	Nobel Prize	Ph.D. in Physics	Unified theory of electromagnetism
4997	Charles Darwin	UK	Physics	181	Quantum Theory	1942	Male	Bohr Model	Two Nobel Prizes	Ph.D.	Contributions to cosmology and quantum gravity
4998	Stephen Hawking	Austria	Mathematics	130	Cosmos Series	1975	Female	Cosmos	Nobel Prize	Ph.D. in Physics	Inventor and electrical engineering pioneer
4999	Niels Bohr	USA	Physics	184	Wave Equation	1969	Female	Vitruvian Man	Nobel Prize	Ph.D. in Mathematics	Unified theory of electromagnetism

shape

It show the total no of rows & Column in the dataset

```
In [21]: df.shape
```

```
Out[21]: (5000, 11)
```

Columns

It show the no of each Column

```
In [25]: df.columns

Out[25]: Index(['Name', 'Country', 'Field of Expertise', 'IQ', 'Achievements', 'Birth Year', 'Gender', 'Notable Works', 'Awards', 'Education', 'Influence'],
              dtype='object')
```

.dtypes

This Attribute show the data type of each column

```
In [29]: df.dtypes

Out[29]: Name                object
Country              object
Field of Expertise    object
IQ                   int64
Achievements         object
Birth Year            int64
Gender               object
Notable Works        object
Awards               object
Education             object
Influence            object
dtype: object
```

.unique()

In a column, It show the unique value of specific column.

```
In [33]: df["Country"].unique()

Out[33]: array(['Austria', 'Italy', 'UK', 'USA', 'Germany', 'Denmark', 'France'],
              dtype=object)
```

.nunique()

It will show the total no of unique value from whole data frame

```
In [37]: df.nunique()

Out[37]: Name                26
Country              7
Field of Expertise    7
IQ                   71
Achievements         19
Birth Year            181
Gender               12
Notable Works        26
Awards               7
Education             9
Influence            19
dtype: int64
```

.describe()

It show the Count, mean , median etc

```
In [49]: df.describe()

Out[41]:
```

	IQ	Birth Year
count	5000.000000	5000.000000
mean	164.765200	1949.527400
std	20.524169	29.024806
min	130.000000	1900.000000
25%	147.000000	1924.000000
50%	165.000000	1950.000000
75%	182.000000	1974.000000
max	200.000000	2000.000000

.value_counts

It Shows all the unique values with their count

```
In [45]: df["Country"].value_counts()

Out[45]: Country    2335
USA                987
Italy              786
Denmark           257
France            249
Germany           245
Austria           241
Name: count, dtype: int64
```

isnull()

It shows the how many null values

```
In [49]: df.isnull()

Out[49]:
```

	Name	Country	Field of Expertise	IQ	Achievements	Birth Year	Gender	Notable Works	Awards	Education	Influence
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...
4995	False	False	False	False	False	False	False	False	True	False	False
4996	False	False	False	False	False	False	False	False	False	False	False
4997	False	False	False	False	False	False	False	False	False	False	False
4998	False	False	False	False	False	False	False	False	False	False	False
4999	False	False	False	False	False	False	False	False	False	False	False

5000 rows × 11 columns

.info()

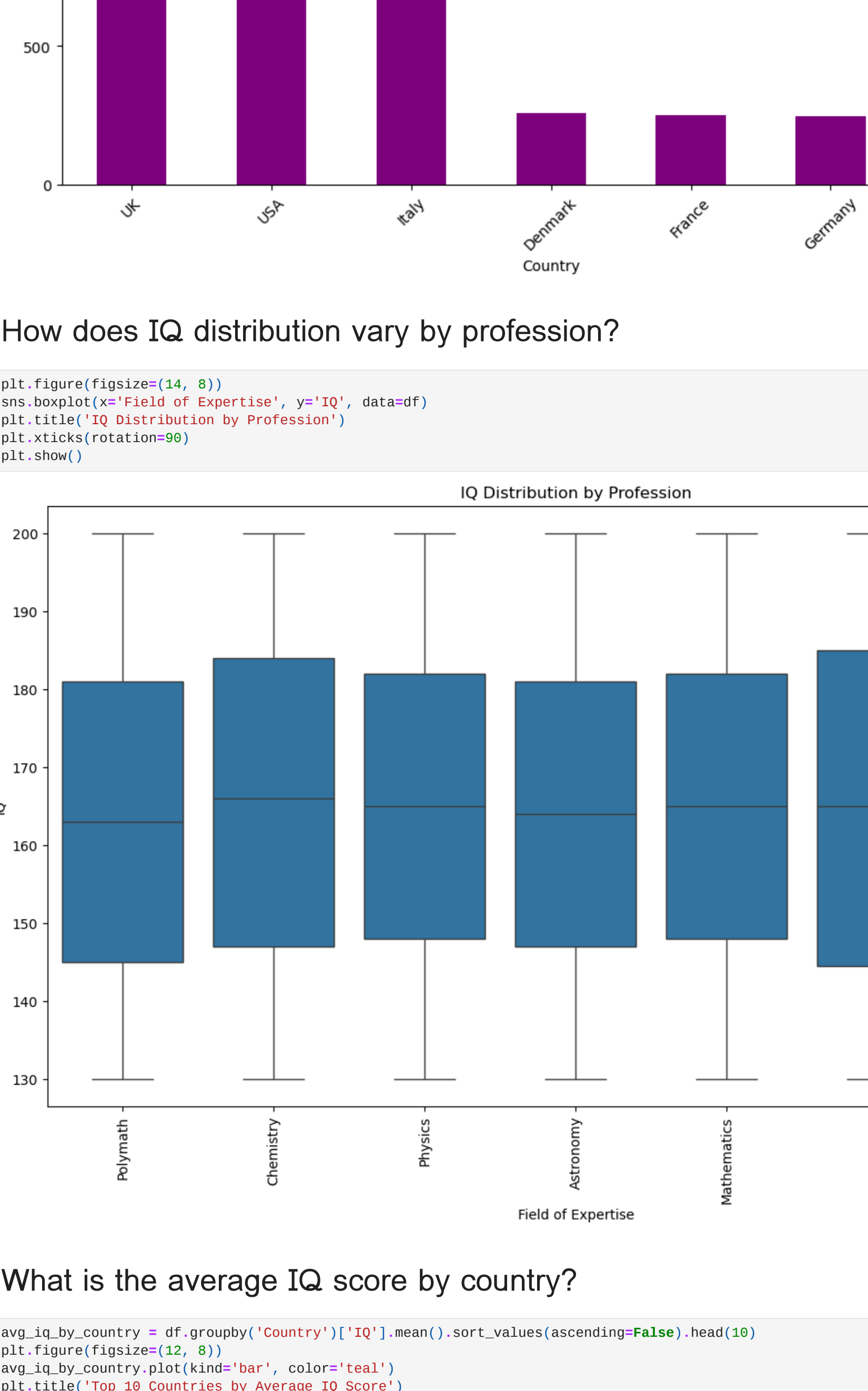
To Show Data type of each column

```
In [53]: df.info()

Out[53]:
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5000 entries, 0 to 4999
Data columns (total 11 columns):
 #   column                non-null count  dtype
---  ---
 0   Name                  5000 non-null object
 1   Country              5000 non-null object
 2   Field of Expertise    5000 non-null object
 3   IQ                   5000 non-null int64
 4   Achievements         5000 non-null object
 5   Birth Year            5000 non-null int64
 6   Gender               5000 non-null object
 7   Notable Works        5000 non-null object
 8   Awards               3753 non-null object
 9   Education             5000 non-null object
10  Influence            5000 non-null object
dtypes: int64(2), object(9)
memory usage: 429.8+ KB
```

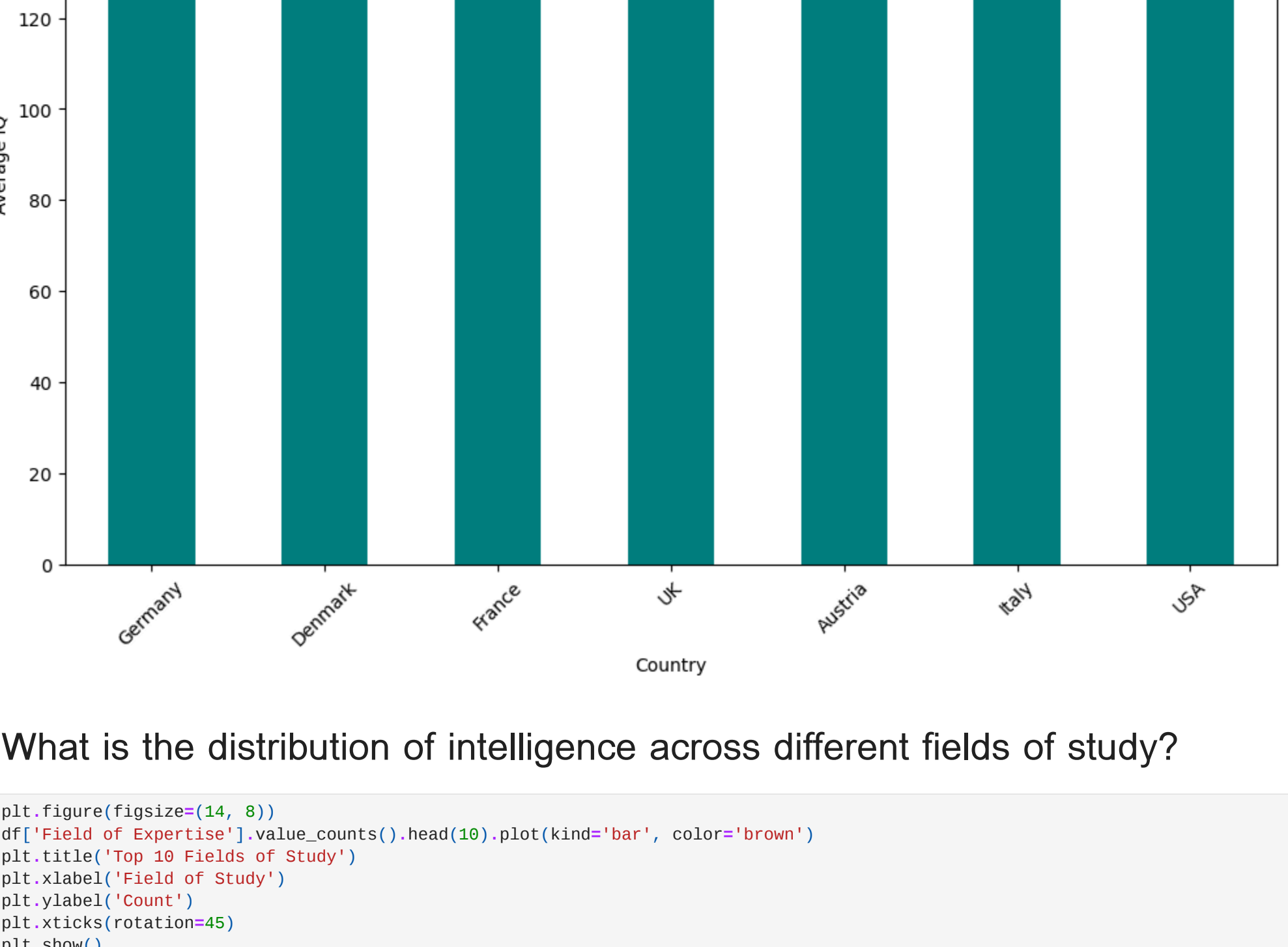
What is the distribution of IQ scores among the top intelligent people?

```
In [57]: plt.figure(figsize=(10, 6))
sns.histplot(df['IQ'], kde=True, color='blue')
plt.title('Distribution of IQ Scores')
plt.xlabel('IQ Score')
plt.ylabel('Frequency')
plt.show()
```



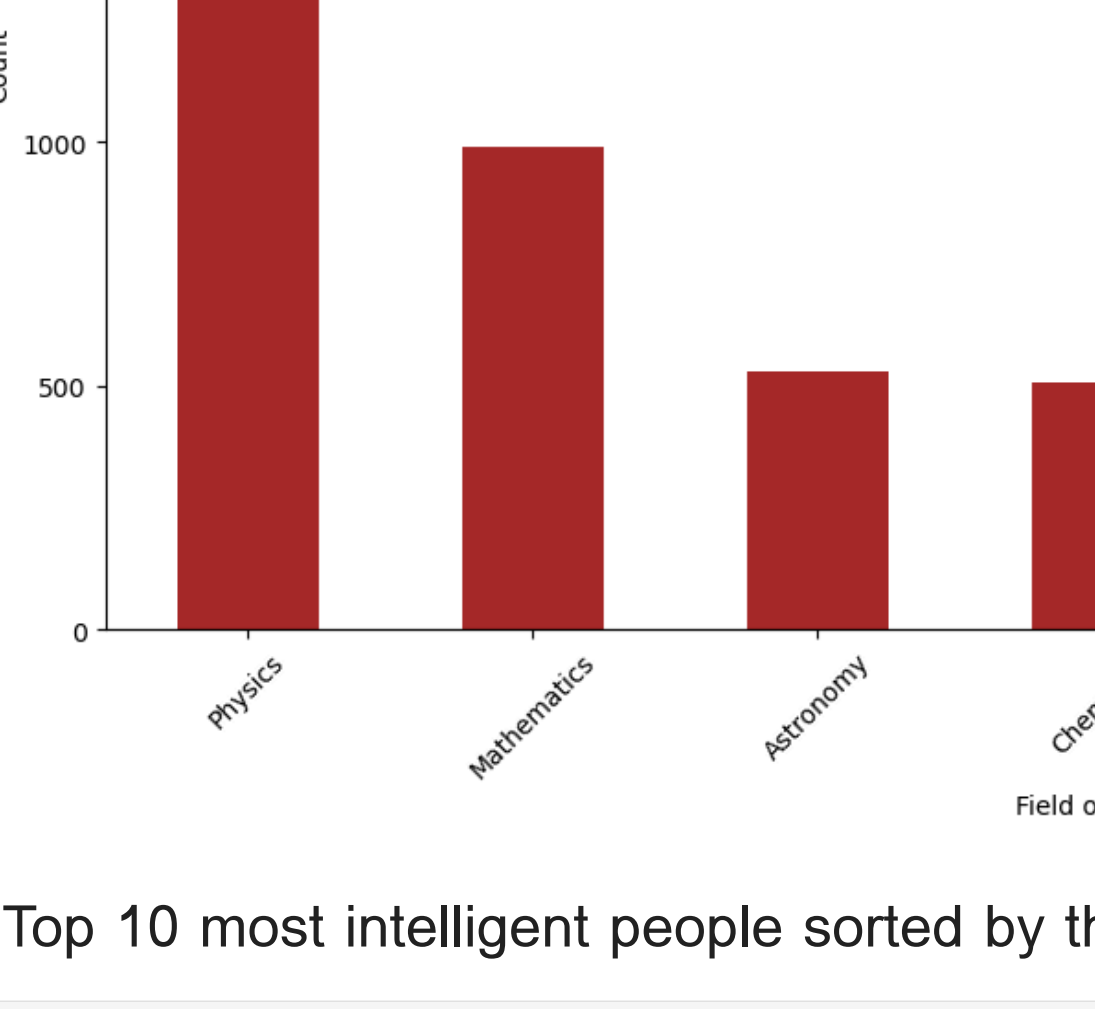
Which professions are most common among the top intelligent people?

```
In [62]: plt.figure(figsize=(12, 8))
df['Field of Expertise'].value_counts().head(10).plot(kind='bar', color='green')
plt.title('Top 10 Most Common Professions')
plt.xlabel('Field of Expertise')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



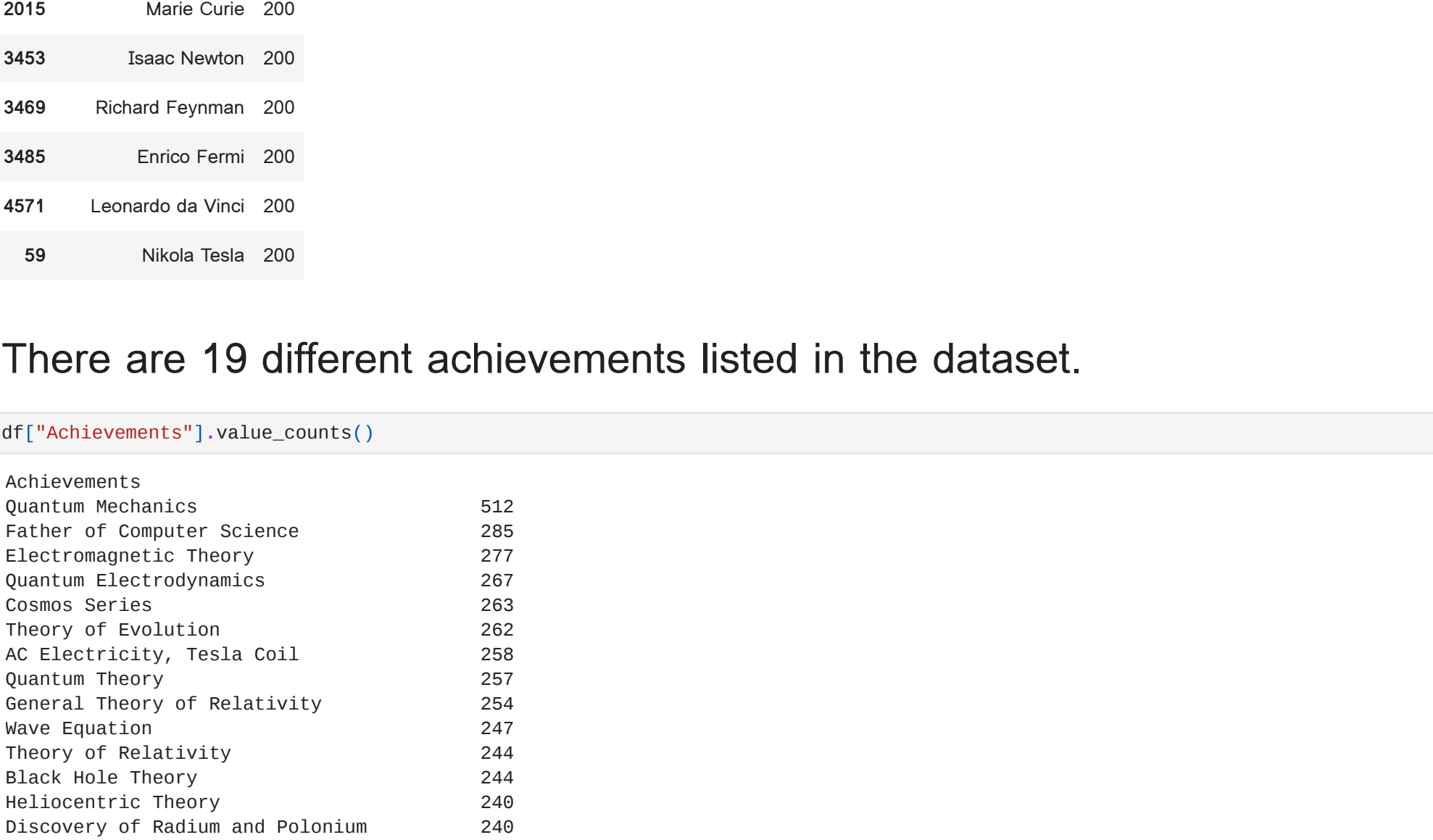
What is the gender distribution among the top intelligent people?

```
In [65]: gender_counts = df['Gender'].value_counts()
sns.histplot(gender_counts.index, autopct='%1.1f%%', startangle=140, color=['skyblue', 'pink'])
plt.title('Gender Distribution')
plt.show()
```



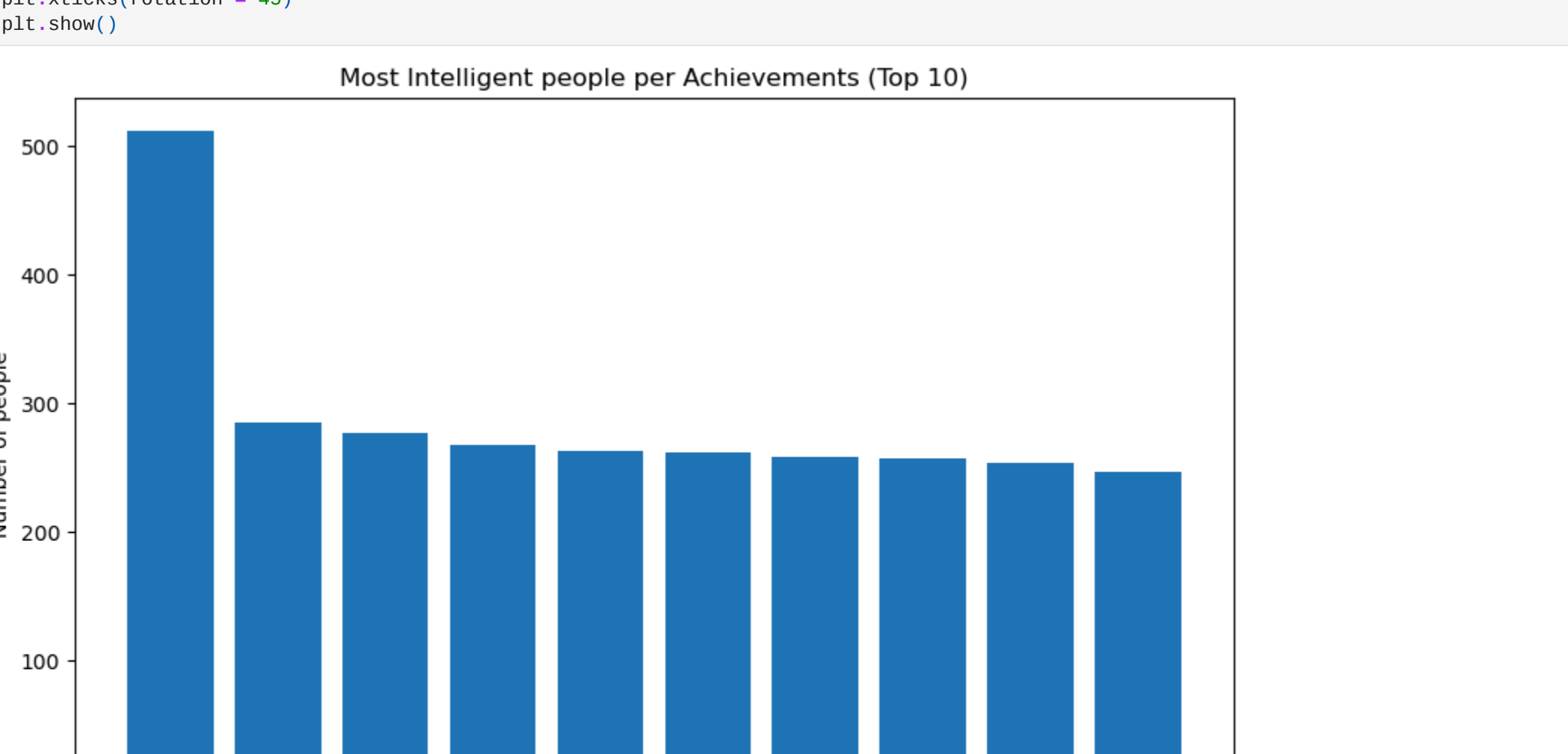
What are the top 10 countries with the highest number of intelligent people?

```
In [70]: plt.figure(figsize=(12, 8))
df['Country'].value_counts().head(10).plot(kind='bar', color='purple')
plt.title('Top 10 Countries with Most Intelligent People')
plt.xlabel('Country')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



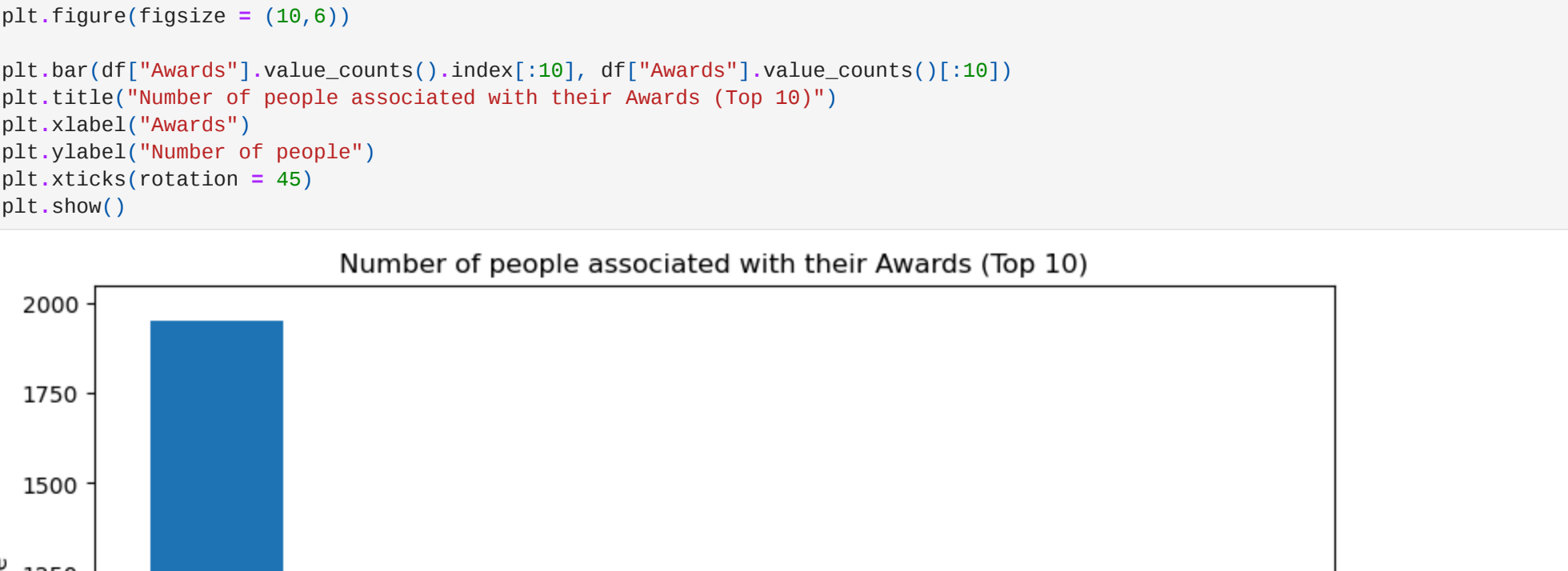
How does IQ distribution vary by profession?

```
In [76]: plt.figure(figsize=(14, 8))
sns.boxplot(x='Field of Expertise', y='IQ', data=df)
plt.title('IQ Distribution by Profession')
plt.show()
```



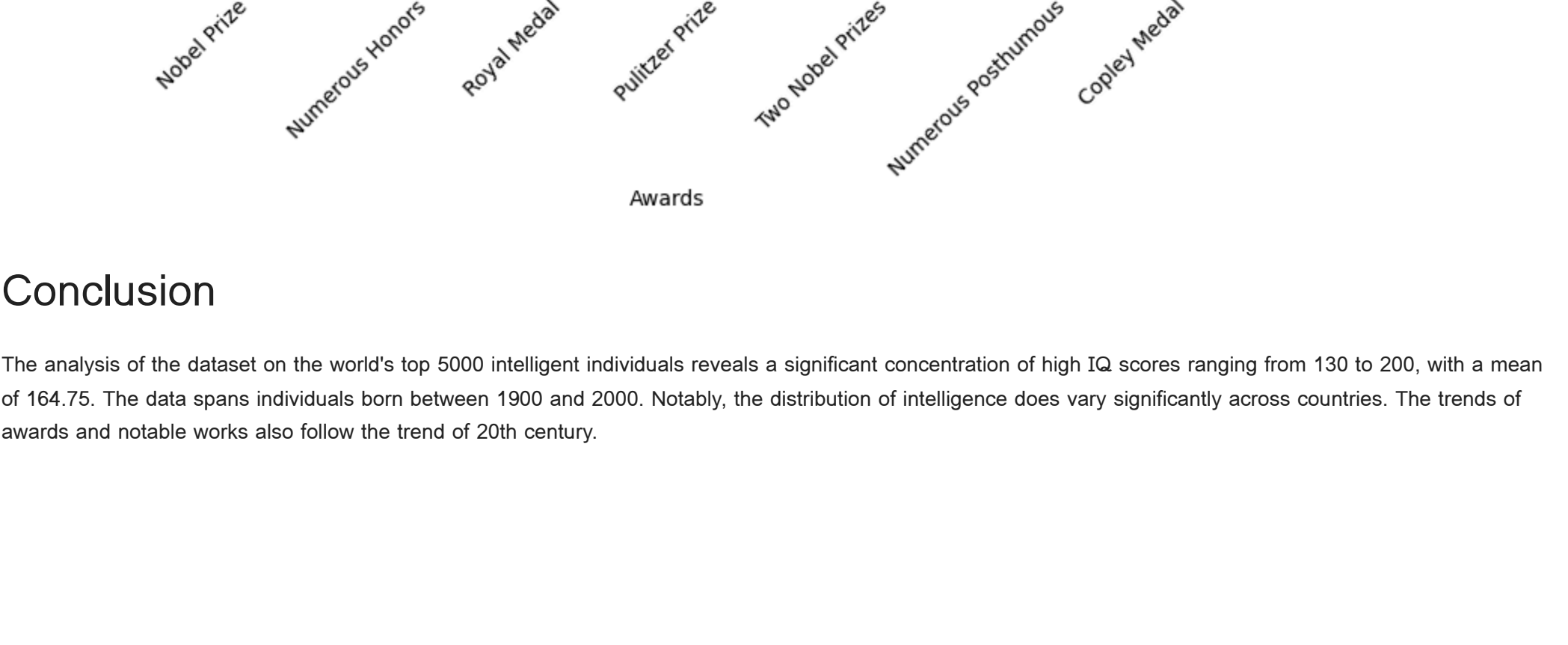
What is the average IQ score by country?

```
In [82]: avg_iq_by_country = df.groupby('Country')['IQ'].mean().sort_values(ascending=False).head(10)
df['Field of Expertise'].value_counts().head(10).plot(kind='bar', color='teal')
plt.title('Top 10 Countries by Average IQ Score')
plt.xlabel('Country')
plt.ylabel('Average IQ')
plt.xticks(rotation=45)
plt.show()
```



What is the distribution of intelligence across different fields of study?

```
In [87]: plt.figure(figsize=(14, 8))
df['Field of Expertise'].value_counts().head(10).plot(kind='bar', color='brown')
plt.title('Top 10 Fields of Study')
plt.xlabel('Field of Study')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```



Top 10 most intelligent people sorted by their IQ

```
In [92]: iq_top_10 = pd.DataFrame()
iq_top_10['Name'] = df.sort_values(by='IQ', ascending=False)['Name'][:10]
iq_top_10['IQ'] = df.sort_values(by='IQ', ascending=False)['IQ'][:10]
iq_top_10
```

```
Out[92]:
```

	Name	IQ
3585	Roger Penrose	200
1876	Michael Faraday	200
2027	Roger Penrose	200
4828	James Clerk Maxwell	200
2015	Maria Curie	200
3453	Isaac Newton	200
3469	Richard Feynman	200
3485	Enrico Fermi	200
4571	Leonardo da Vinci	200
59	Nikola Tesla	200

There are 19 different achievements listed in the dataset.

```
In [95]: df['Achievements'].value_counts()

Out[95]: Achievements
Quantum Mechanics          512
Father of Computer Science 285
Electromagnetic Theory     277
Quantum Electrodynamics    267
Cosmos Series              263
Theory of Evolution        262
AC Electricity, Tesla Coil 258
Quantum Theory             257
General Theory of Relativity 247
Wave Equation              244
Black Hole Theory          244
Heliocentric Theory        240
Discovery of Radium and Polonium 236
First computer algorithm    236
Electromagnetic Induction   235
Theory of Motion            229
Nuclear Reactor            226
Mona Lisa, The Last Supper, Inventions 225
Name: count, dtype: int64
```

Most Intelligent people per Achievements (Top 10)

```
In [108]: plt.figure(figsize=(10, 6))
plt.bar(df['Achievements'].value_counts().index[:10], df['Achievements'].value_counts()[:10])
plt.title('Most Intelligent people per Achievements (Top 10)')
plt.xlabel('Achievements')
plt.ylabel('Number of people')
plt.xticks(rotation=45)
plt.show()
```


Top 10 Number of people associated with their Awards

```
In [105]: plt.figure(figsize=(10, 6))
plt.bar(df['Awards'].value_counts().index[:10], df['Awards'].value_counts()[:10])
plt.title('Number of people associated with their Awards (Top 10)')
plt.xlabel('Awards')
plt.ylabel('Number of people')
plt.xticks(rotation=45)
plt.show()
```


Conclusion

The analysis of the dataset on the world's top 5000 intelligent individuals reveals a significant concentration of high IQ scores ranging from 130 to 200, with a mean of 164.75. The data spans individuals born between 1900 and 2000. Notably, the distribution of intelligence does vary significantly across countries. The trends of awards and notable works also follow the trend of 20th century.