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


data_april = pd.read_csv("/content/drive/MyDrive/Machine learning (personal)/Data/VoiceCalling/CallVoiceQualityExperience-2018-April.csv
data_may = pd.read_csv("/content/drive/MyDrive/Machine learning (personal)/Data/VoiceCalling/CallVoiceQuality_Data_2018_May.csv")


# Add a 'Month' column to differentiate between April and May data
data_april['Month'] = 'April'
data_may['Month'] = 'May'


# Concatenate the data
merged_data = pd.concat([data_april, data_may], ignore_index=True)


# Convert Rating to numeric
merged_data['Rating'] = pd.to_numeric(merged_data['Rating'], errors='coerce')



# Vertical bar chart of average call quality rate per operator
plt.figure(figsize=(12, 6))
operator_avg_rating = merged_data.groupby(['Operator', 'Month'])['Rating'].mean().unstack().sort_values(by='May', ascending=False)
operator_avg_rating.plot(kind='bar', ax=plt.gca())
plt.title('Average Call Quality Rate per Operator')
plt.xlabel('Operator')
plt.ylabel('Average Rating')
plt.xticks(rotation=45)
plt.legend(title='Month')
plt.show()
```
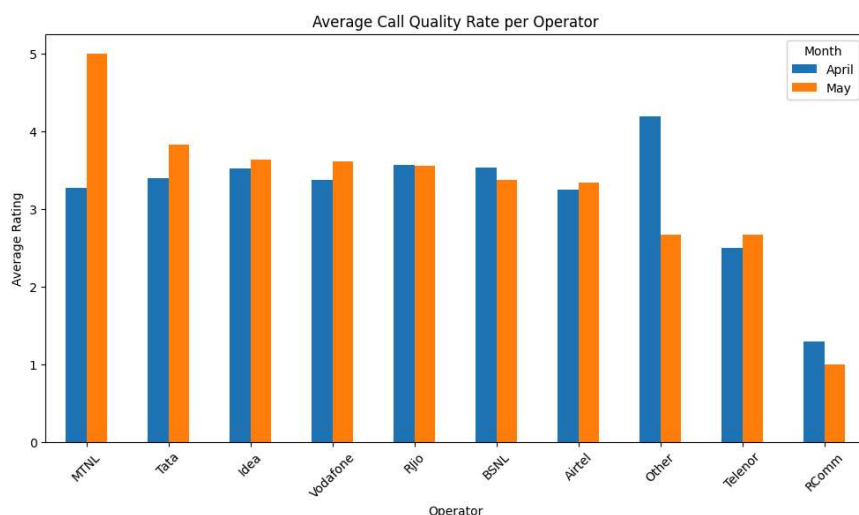


```python
# Vertical bar chart of the quality level per each state in India
plt.figure(figsize=(12, 6))
state_avg_rating = merged_data.groupby(['State Name', 'Month'])['Rating'].mean().unstack().sort_values(by='May', ascending=False)
state_avg_rating.plot(kind='bar', ax=plt.gca())
plt.title('Average Call Quality Rate per State')
plt.xlabel('State')
plt.ylabel('Average Rating')
plt.xticks(rotation=90)
plt.legend(title='Month')
plt.show()
```

Average Call Quality Rate per State
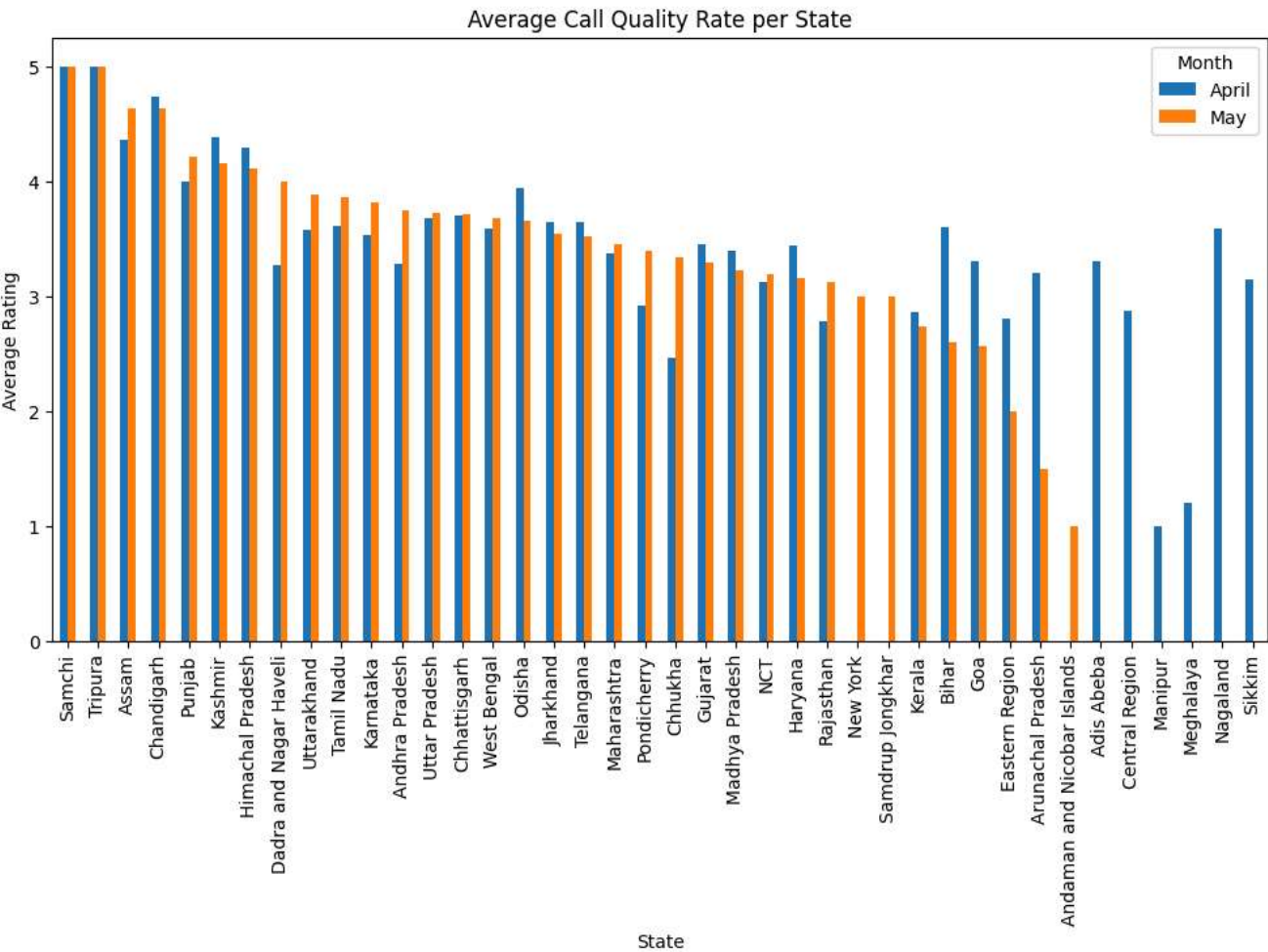
```
# Vertical bar chart showing the relationship between the call quality and the network type
plt.figure(figsize=(10, 6))
sns.barplot(x='Network Type', y='Rating', hue='Month', data=merged_data, order=['2G', '3G', '4G'])
plt.title('Average Call Quality Rate per Network Type')
plt.xlabel('Network Type')
plt.ylabel('Average Rating')
plt.legend(title='Month')
plt.show()
```
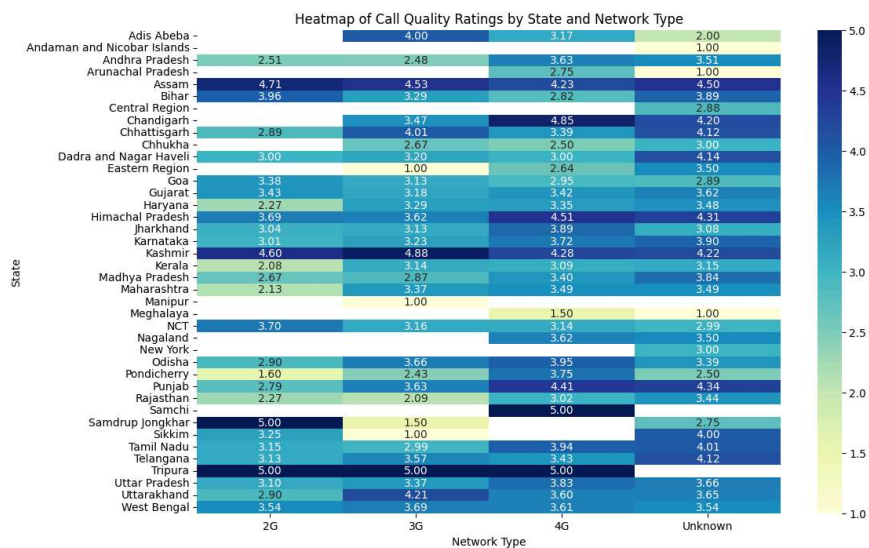

Average Call Quality Rate per Network Type
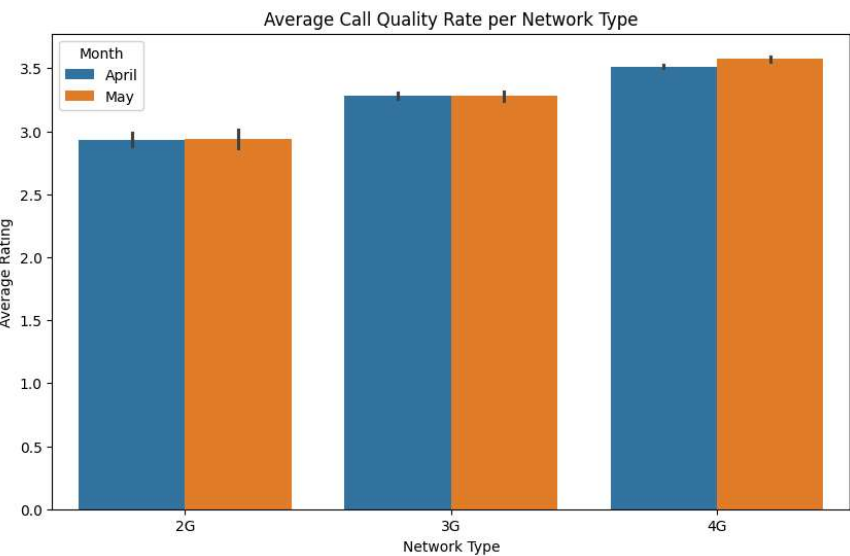
```
# Horizontal bar chart of average call quality rate per Call Drop Category
plt.figure(figsize=(10, 6))
call_drop_avg_rating = merged_data.groupby(['Call Drop Category', 'Month'])['Rating'].mean().unstack().sort_values(by='May', ascending=F
call_drop_avg_rating.plot(kind='barh', ax=plt.gca())
plt.title('Average Call Quality Rate per Call Drop Category')
plt.xlabel('Average Rating')
plt.ylabel('Call Drop Category')
plt.legend(title='Month')
plt.show()
```
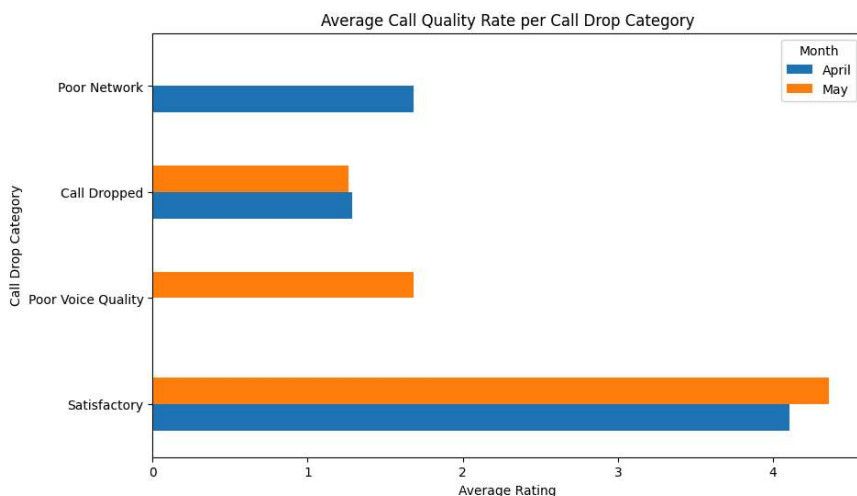


```
# Heat map between state, Network Type and rating
state_network_pivot = merged_data.pivot_table(index='State Name', columns='Network Type', values='Rating', aggfunc='mean')
plt.figure(figsize=(12, 8))
sns.heatmap(state_network_pivot, cmap='YlGnBu', annot=True, fmt=".2f")
plt.title('Heatmap of Call Quality Ratings by State and Network Type')
plt.xlabel('Network Type')
plt.ylabel('State')
plt.show()
```

Heatmap of Call Quality Ratings by State and Network Type

State	2G	3G	4G	Unknown
Adis Abeba		4.00	3.17	2.00
Andaman and Nicobar Islands				1.00
Andhra Pradesh	2.51	2.48	3.63	3.51
Arunachal Pradesh			2.75	1.00
Assam	4.71	4.53	4.23	4.50
Bihar	3.96	3.29	2.82	3.89
Central Region				2.88
Chandigarh		3.47	4.85	4.20
Chhattisgarh	2.89	4.01	3.39	4.12
Chhukha		2.67	2.50	3.00
Dadra and Nagar Haveli	3.00	3.20	3.00	4.14
Eastern Region		1.00	2.64	3.50
Goa	3.38	3.13	2.95	2.89
Gujarat	3.43	3.18	3.42	3.62
Haryana	2.27	3.29	3.35	3.48
Himachal Pradesh	3.69	3.62	4.51	4.31
Jharkhand	3.04	3.13	3.89	3.08
Karnataka	3.01	3.23	3.72	3.90
Kashmir	4.60	4.88	4.28	4.22
Kerala	2.08	3.14	3.09	3.15
Madhya Pradesh	2.67	2.87	3.40	3.84
Maharashtra	2.13	3.37	3.49	3.49
Manipur		1.00		
Meghalaya			1.50	1.00
NCT	3.70	3.16	3.14	2.99
Nagaland			3.62	3.50
New York				3.00
Odisha	2.90	3.66	3.95	3.39
Pondicherry	1.60	2.43	3.75	2.50
Punjab	2.79	3.63	4.41	4.34
Rajasthan	2.27	2.09	3.02	3.44
Samchi			5.00	
Samdrup Jongkhar	5.00	1.50		2.75
Sikkim	3.25	1.00		4.00
Tamil Nadu	3.15	2.99	3.94	4.01
Telangana	3.13	3.57	3.43	4.12
Tripura	5.00	5.00	5.00	
Uttar Pradesh	3.10	3.57	3.83	3.66
Uttarakhand	2.90	4.21	3.60	3.65
West Bengal	3.54	3.69	3.61	3.54

Network Type

```
# Vertical bar chart of average call quality rate per Indoor_Outdoor_Travelling
plt.figure(figsize=(10, 6))
io_travelling_avg_rating = merged_data.groupby('In Out Travelling')['Rating'].mean().sort_values(ascending=False)
sns.barplot(x=io_travelling_avg_rating.index, y=io_travelling_avg_rating.values)
plt.title('Average Call Quality Rate per Indoor_Outdoor_Travelling')
plt.xlabel('Indoor_Outdoor_Travelling')
plt.ylabel('Average Rating')
plt.show()
```


Average Call Quality Rate per Indoor_Outdoor_Travelling
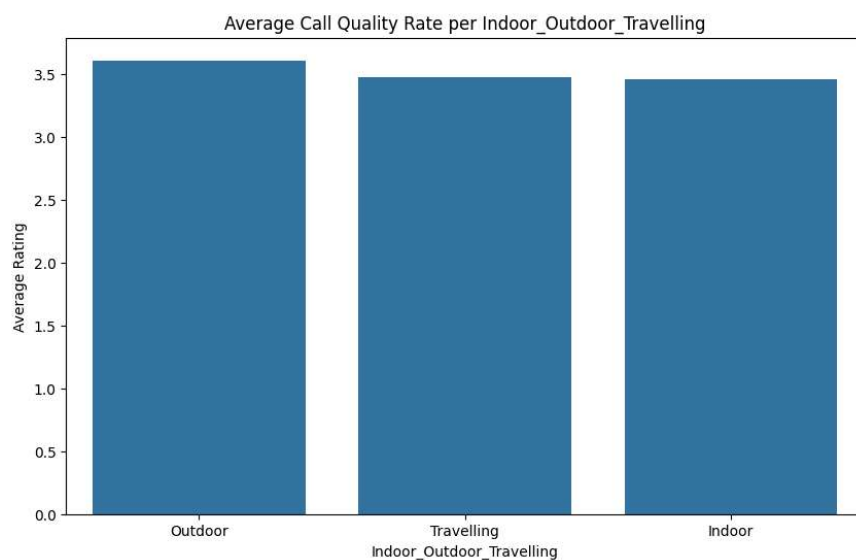
#1. What is the level of overall call quality? (by the operator, by the network.

```
# Average call quality by operator
operator_avg_quality = merged_data.groupby('Operator')['Rating'].mean().sort_values(ascending=False)
print(operator_avg_quality)
```

```
    Operator
    Other       3.888889
    Tata        3.624703
    RJio        3.562902
    Idea        3.560112
    BSNL        3.485532
    Vodafone    3.451958
    MTNL        3.291667
    Airtel      3.277418
    Telenor     2.659574
    RComm       1.287671
    Name: Rating, dtype: float64
```

```
# Average call quality by network type
network_avg_quality = merged_data.groupby('Network Type')['Rating'].mean().sort_values(ascending=False)
print(network_avg_quality)
```

```
    Network Type
    Unknown    3.615250
    4G         3.535992
    3G         3.282701
    2G         2.935904
    Name: Rating, dtype: float64
```

#2. Which operators provide low-value services?

```
# Operators with lowest average call quality ratings
low_value_operators = operator_avg_quality.nsmallest(5)  # Selecting bottom 5 operators
print(low_value_operators)
```