

```
In [3]: import os
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
# Убедитесь, что phik установлен. Если нет, выполните: pip install phik
from phik import resources, report

# Загрузка данных
current_dir = os.getcwd()
file_path = os.path.join(current_dir, 'churn.csv')
df = pd.read_csv(file_path)

# Предобработка данных
# Исключаем customerID, поскольку он уникальный для каждого пользователя
df.drop('customerID', axis=1, inplace=True)

# Кодирование категориальных переменных
for col in df.select_dtypes(include=['object']).columns:
    if col != 'TotalCharges': # Исключаем TotalCharges из кодирования
        df[col], _ = pd.factorize(df[col])

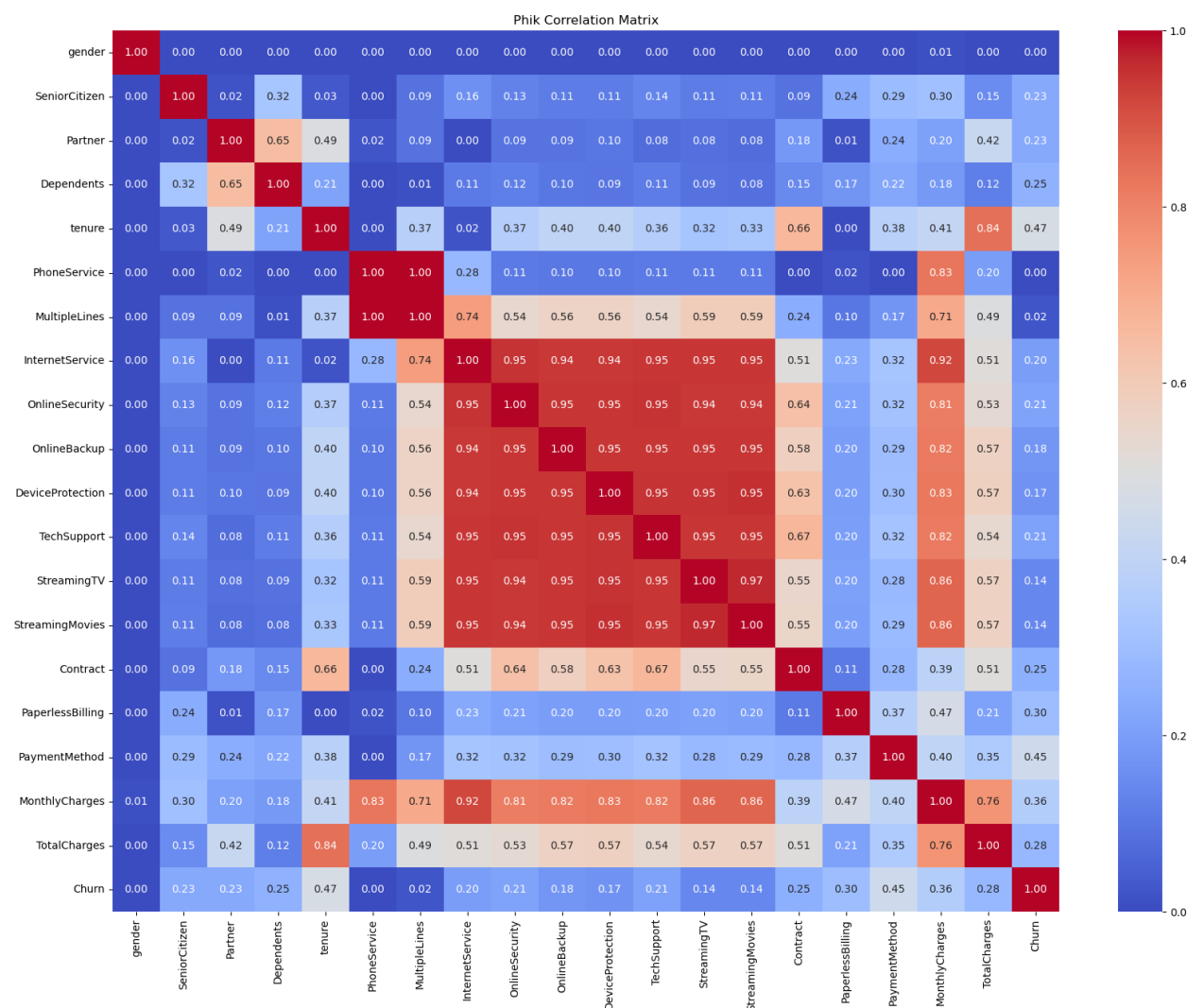
# Преобразование TotalCharges из строки в числовой формат и обработка пус
df['TotalCharges'] = pd.to_numeric(df['TotalCharges'], errors='coerce')
df['TotalCharges'].fillna(0, inplace=True)

# Анализ взаимосвязей

# Определение колонок, которые следует рассматривать как интервальные
interval_cols = ['tenure', 'MonthlyCharges', 'TotalCharges']

# Вычисление корреляции Phik
phik_overview = df.phik_matrix(interval_cols=interval_cols)

# Визуализация через тепловую карту
plt.figure(figsize=(20, 15))
sns.heatmap(phik_overview, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Phik Correlation Matrix")
plt.show()
```



In []: