In [1]:
```
# -*- coding: utf-8 -*-
"""final dissertation file .ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1QnyW4g8FmOniu78SxlxFhx9iI5T3Bc4v
"""

# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
# Load the dataset
data = pd.read_csv("payment-practices.csv")
```

In [4]:
```
# Step 1: Evaluate the state of supply chain management in the UK
# Calculate some basic statistics
mean_time_to_pay = data["Average time to pay"].mean()
percentage_paid_within_30_days = data["% Invoices paid within 30 days"].mean()
percentage_paid_between_31_and_60_days = data["% Invoices paid between 31 and 60 days"
percentage_paid_later_than_60_days = data["% Invoices paid later than 60 days"].mean()
```

In [5]:
```
# Display the results
print("Mean Time to Pay:", mean_time_to_pay)
print("Percentage Paid within 30 Days:", percentage_paid_within_30_days)
print("Percentage Paid between 31 and 60 Days:", percentage_paid_between_31_and_60_day
print("Percentage Paid Later than 60 Days:", percentage_paid_later_than_60_days)
```

```
Mean Time to Pay: 37.26472884235662
Percentage Paid within 30 Days: 53.93722734995637
Percentage Paid between 31 and 60 Days: 31.614685218349635
Percentage Paid Later than 60 Days: 14.43871975019516
```

In [6]:
```
# Step 2: Assess how the UK supply chain is implementing Industry 4.0 technologies
# Count the number of companies offering E-Invoicing and Supply-chain financing
e_invoicing_count = data["E-Invoicing offered"].sum()
supply_chain_financing_count = data["Supply-chain financing offered"].sum()
```

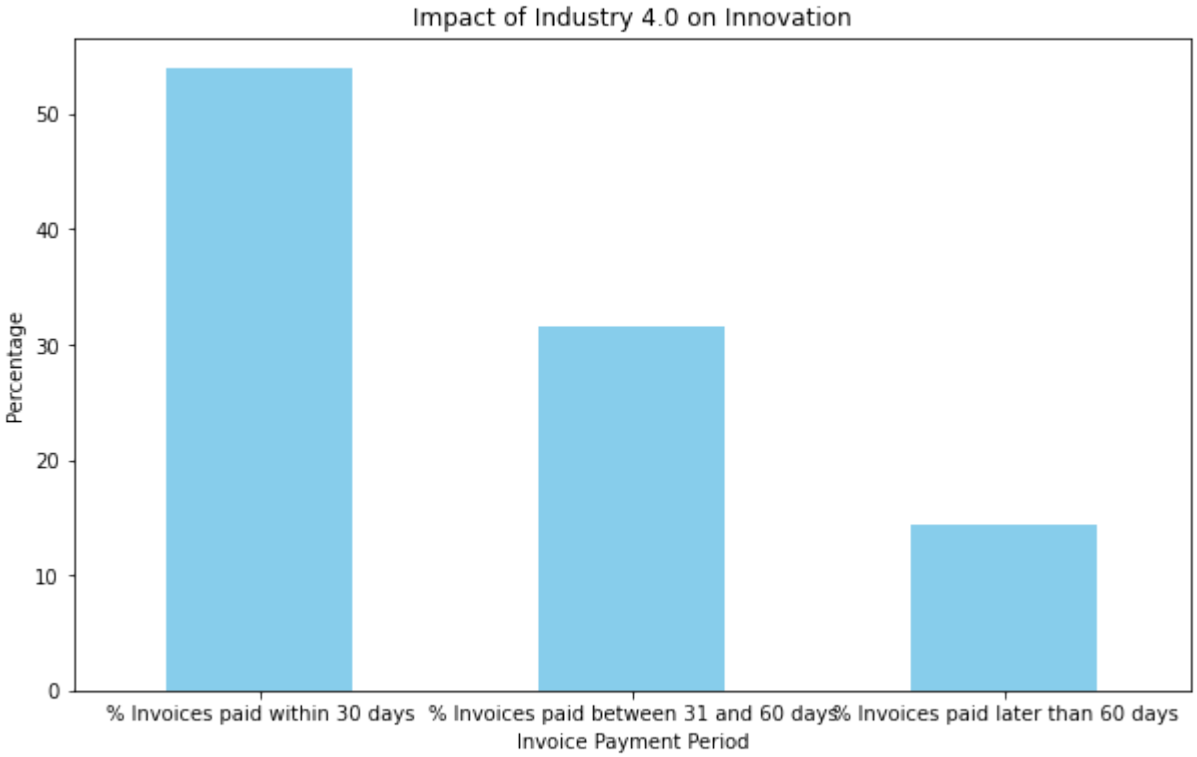
In [7]:
```
# Step 2: Assess how the UK supply chain is implementing Industry 4.0 technologies
# Count the number of companies offering E-Invoicing and Supply-chain financing
e_invoicing_count = data["E-Invoicing offered"].sum()
supply_chain_financing_count = data["Supply-chain financing offered"].sum()
```

In [8]:
```
# Display the results
print("Number of Companies Offering E-Invoicing:", e_invoicing_count)
print("Number of Companies Offering Supply-chain Financing:", supply_chain_financing_c
```

```
Number of Companies Offering E-Invoicing: 5210
Number of Companies Offering Supply-chain Financing: 1467
```

In [9]:
```
# Step 3: Visualize the impact of Industry 4.0 on innovation and efficiency
# Calculate the mean values for innovation-related columns
innovation_columns = ["% Invoices paid within 30 days", "% Invoices paid between 31 ar
innovation_means = data[innovation_columns].mean()
```

In [10]:
```
# Plot the results
plt.figure(figsize=(10, 6))
innovation_means.plot(kind='bar', color='skyblue')
plt.title('Impact of Industry 4.0 on Innovation')
plt.xlabel('Invoice Payment Period')
plt.ylabel('Percentage')
plt.xticks(rotation=0)
plt.show()
```



In [11]:
```
# Step 4: Investigate how partnerships and collaboration optimize supply chains
# Count the number of companies participating in payment codes
payment_codes_count = data["Participates in payment codes"].sum()
```

In [12]:
```
# Display the results
print("Number of Companies Participating in Payment Codes:", payment_codes_count)
```

```
Number of Companies Participating in Payment Codes: 2186
```

In [13]:
```
# 5. Overview of the dataset
print("Dataset Overview:")
print(data.info())
```

```
Dataset Overview:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 24225 entries, 0 to 24224
Data columns (total 23 columns):
 #   Column                                                    Non-Null Count  Dtype
---  ------                                                    --------------  -----
 0   Report Id                                                 24225 non-null  int64
 1   Start date                                                24225 non-null  object
 2   End date                                                  24225 non-null  object
 3   Filing date                                               24225 non-null  object
 4   Company                                                   24225 non-null  object
 5   Company number                                            24225 non-null  object
 6   Payments made in the reporting period                     13844 non-null  object
 7   Average time to pay                                       21777 non-null  float64
 8   % Invoices paid within 30 days                            21777 non-null  float64
 9   % Invoices paid between 31 and 60 days                    21777 non-null  float64
 10  % Invoices paid later than 60 days                        21777 non-null  float64
 11  % Invoices not paid within agreed terms                   21820 non-null  float64
 12  Shortest (or only) standard payment period                21820 non-null  float64
 13  Longest standard payment period                           14306 non-null  float64
 14  Maximum contractual payment period                        21820 non-null  float64
 15  Payment terms have changed                                21820 non-null  object
 16  Suppliers notified of changes                             541 non-null    object
 17  Participates in payment codes                             24225 non-null  bool
 18  E-Invoicing offered                                       21820 non-null  object
 19  Supply-chain financing offered                            21820 non-null  object
 20  Policy covers charges for remaining on supplier list      21820 non-null  object
 21  Charges have been made for remaining on supplier list     21820 non-null  object
 22  URL                                                       24225 non-null  object
dtypes: bool(1), float64(8), int64(1), object(13)
memory usage: 4.1+ MB
None
```

In [14]:
```python
# 6. Summary statistics
print("\nSummary Statistics:")
print(data.describe())
```

Summary Statistics:

	Report Id	Average time to pay	% Invoices paid within 30 days
count	24225.000000	21777.000000	21777.000000
mean	12436.194716	37.264729	53.937227
std	7163.162769	25.752755	28.367552
min	2.000000	0.000000	0.000000
25%	6233.000000	25.000000	30.000000
50%	12437.000000	35.000000	55.000000
75%	18642.000000	46.000000	78.000000
max	24815.000000	1000.000000	100.000000

	% Invoices paid between 31 and 60 days
count	21777.000000
mean	31.614685
std	20.689184
min	0.000000
25%	15.000000
50%	30.000000
75%	46.000000
max	100.000000

	% Invoices paid later than 60 days
count	21777.000000
mean	14.438720
std	16.928715
min	0.000000
25%	3.000000
50%	8.000000
75%	19.000000
max	100.000000

	% Invoices not paid within agreed terms
count	21820.000000
mean	29.964299
std	24.561164
min	0.000000
25%	10.000000
50%	24.500000
75%	45.000000
max	100.000000

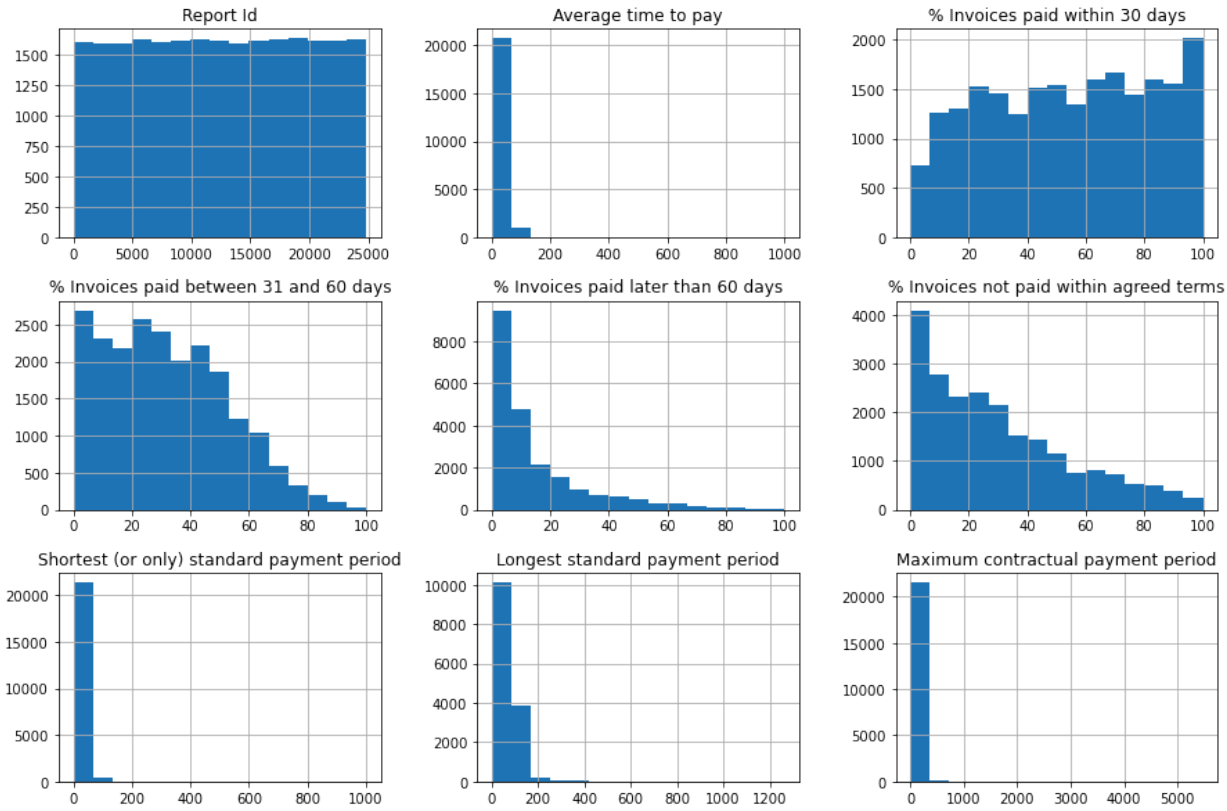
	Shortest (or only) standard payment period
count	21820.000000
mean	21.336709
std	25.709637
min	0.000000
25%	1.000000
50%	20.000000
75%	30.000000
max	1000.000000

	Longest standard payment period	Maximum contractual payment period
count	14306.000000	21820.000000
mean	70.432056	75.190376
std	44.956332	101.612130
min	1.000000	0.000000
25%	60.000000	45.000000
50%	60.000000	60.000000
75%	90.000000	90.000000
max	1264.000000	5475.000000

In [15]:
```
# 7. Check for missing values
print("\nMissing Values:")
print(data.isnull().sum())
```

```
Missing Values:
Report Id                                                    0
Start date                                                   0
End date                                                     0
Filing date                                                  0
Company                                                      0
Company number                                               0
Payments made in the reporting period                    10381
Average time to pay                                       2448
% Invoices paid within 30 days                            2448
% Invoices paid between 31 and 60 days                    2448
% Invoices paid later than 60 days                        2448
% Invoices not paid within agreed terms                   2405
Shortest (or only) standard payment period                2405
Longest standard payment period                           9919
Maximum contractual payment period                        2405
Payment terms have changed                                2405
Suppliers notified of changes                            23684
Participates in payment codes                                0
E-Invoicing offered                                       2405
Supply-chain financing offered                            2405
Policy covers charges for remaining on supplier list      2405
Charges have been made for remaining on supplier list     2405
URL                                                          0
dtype: int64
```

In [16]:
```
# Plot histograms for numeric columns
numeric_columns = data.select_dtypes(include=['float64', 'int64'])
numeric_columns.hist(bins=15, figsize=(15, 10))
plt.suptitle('Numeric Columns Distribution', x=0.5, y=1.02, fontsize=16)
plt.show()
```

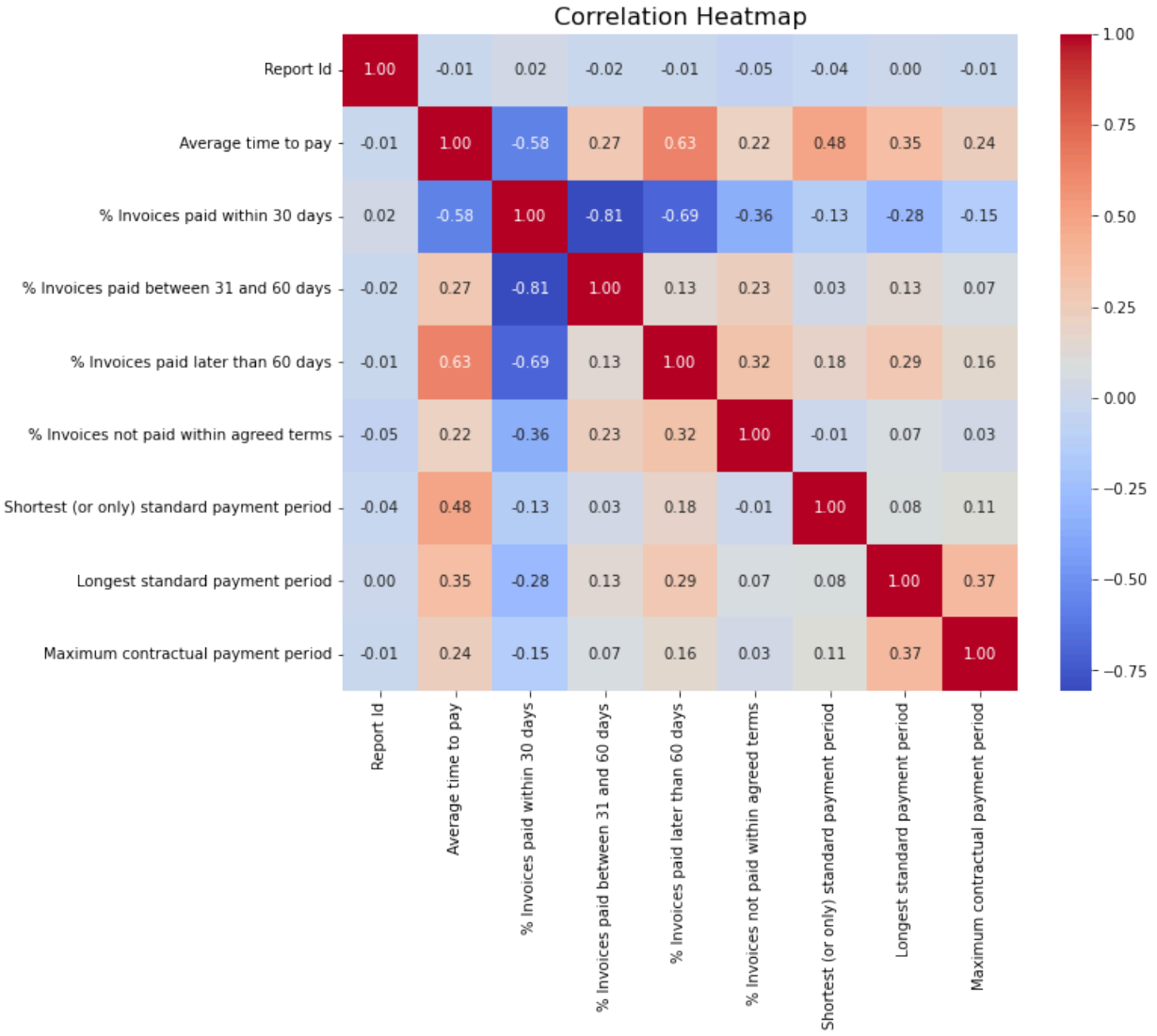


In [17]:
```
# 9. Categorical data analysis

# Count the unique values in each categorical column
categorical_columns = data.select_dtypes(include=['object'])
for column in categorical_columns:
    print(f"\nUnique values in {column}:")
    print(data[column].value_counts())
```

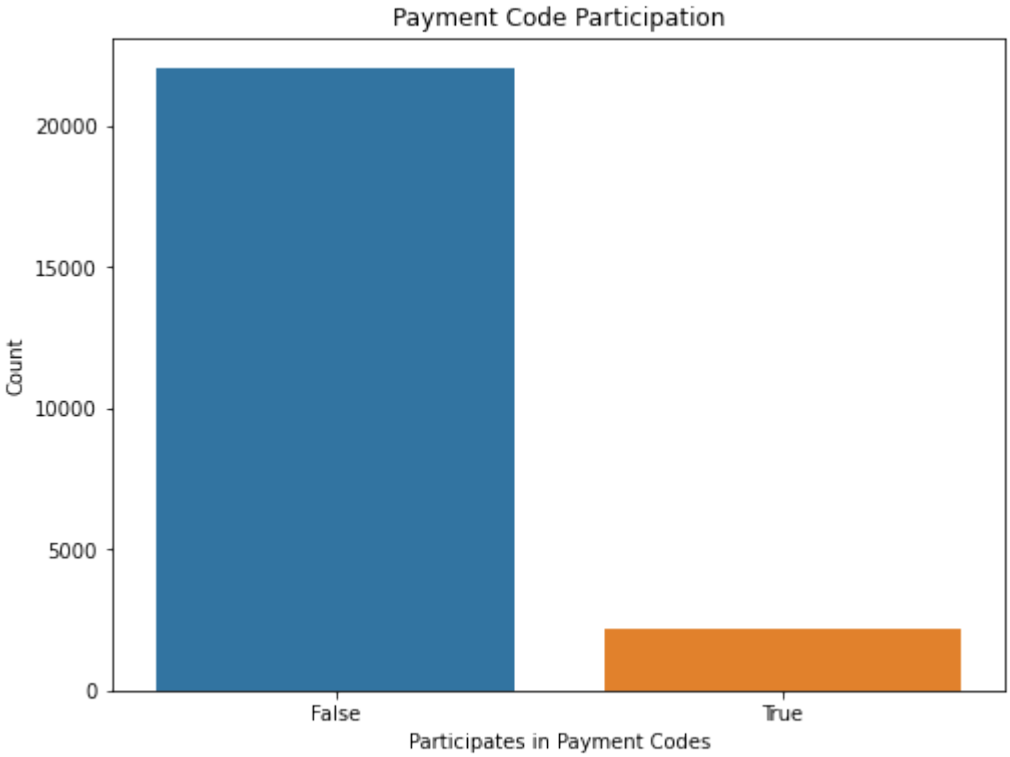
```
Unique values in Start date:
2019-01-01    4355
2018-01-01    4331
2018-07-01    4298
2018-04-01    1483
2018-10-01    1453
              ...
2017-06-28       1
2017-09-16       1
2017-12-24       1
2018-09-27       1
2018-06-24       1
Name: Start date, Length: 301, dtype: int64

Unique values in End date:
2019-06-30    4374
2018-06-30    4310
2018-12-31    4307
2019-03-31    1462
2018-09-30    1451
              ...
2017-12-29       1
2019-04-21       1
2018-08-28       1
2019-04-09       1
2019-01-30       1
Name: End date, Length: 311, dtype: int64

Unique values in Filing date:
2018-07-30    1135
2019-01-30    1052
2019-07-30     957
2018-07-27     668
2019-07-29     617
              ...
2018-08-27       1
2018-08-19       1
2018-05-11       1
2018-05-10       1
2017-11-07       1
Name: Filing date, Length: 601, dtype: int64

Unique values in Company:
MS INTERNATIONAL PLC                              11
THE LANCASTER LANDMARK HOTEL COMPANY LIMITED      9
SIMMONS & SIMMONS LLP                             9
SMITHS MEDICAL INTERNATIONAL LIMITED              7
COMPASS CONTRACT SERVICES (U.K.) LIMITED          7
                                                 ..
WYG ENVIRONMENT PLANNING TRANSPORT LIMITED        1
CAREWATCH CARE SERVICES LIMITED                   1
CAMPBELL LUTYENS & CO. LTD                        1
SERVEST ARTHUR MCKAY LIMITED                      1
LEASEDRIVE LIMITED                                1
Name: Company, Length: 8051, dtype: int64

Unique values in Company number:
00653735    11
OC352713     9
02832349     9
```

```
07465701     7
07875164     7
            ..
10257888     1
04137419     1
10126402     1
02181315     1
01748180     1
Name: Company number, Length: 7863, dtype: int64

Unique values in Payments made in the reporting period:
True     13801
False       43
Name: Payments made in the reporting period, dtype: int64

Unique values in Payment terms have changed:
False    21279
True       541
Name: Payment terms have changed, dtype: int64

Unique values in Suppliers notified of changes:
True     414
False    127
Name: Suppliers notified of changes, dtype: int64

Unique values in E-Invoicing offered:
False    16610
True      5210
Name: E-Invoicing offered, dtype: int64

Unique values in Supply-chain financing offered:
False    20353
True      1467
Name: Supply-chain financing offered, dtype: int64

Unique values in Policy covers charges for remaining on supplier list:
False    21618
True       202
Name: Policy covers charges for remaining on supplier list, dtype: int64

Unique values in Charges have been made for remaining on supplier list:
False    21687
True       133
Name: Charges have been made for remaining on supplier list, dtype: int64

Unique values in URL:
https://check-payment-practices.service.gov.uk/report/2        1
https://check-payment-practices.service.gov.uk/report/16586    1
https://check-payment-practices.service.gov.uk/report/16595    1
https://check-payment-practices.service.gov.uk/report/16594    1
https://check-payment-practices.service.gov.uk/report/16593    1
                                                             ..
https://check-payment-practices.service.gov.uk/report/8315     1
https://check-payment-practices.service.gov.uk/report/8314     1
https://check-payment-practices.service.gov.uk/report/8313     1
https://check-payment-practices.service.gov.uk/report/8312     1
https://check-payment-practices.service.gov.uk/report/24815    1
Name: URL, Length: 24225, dtype: int64
```

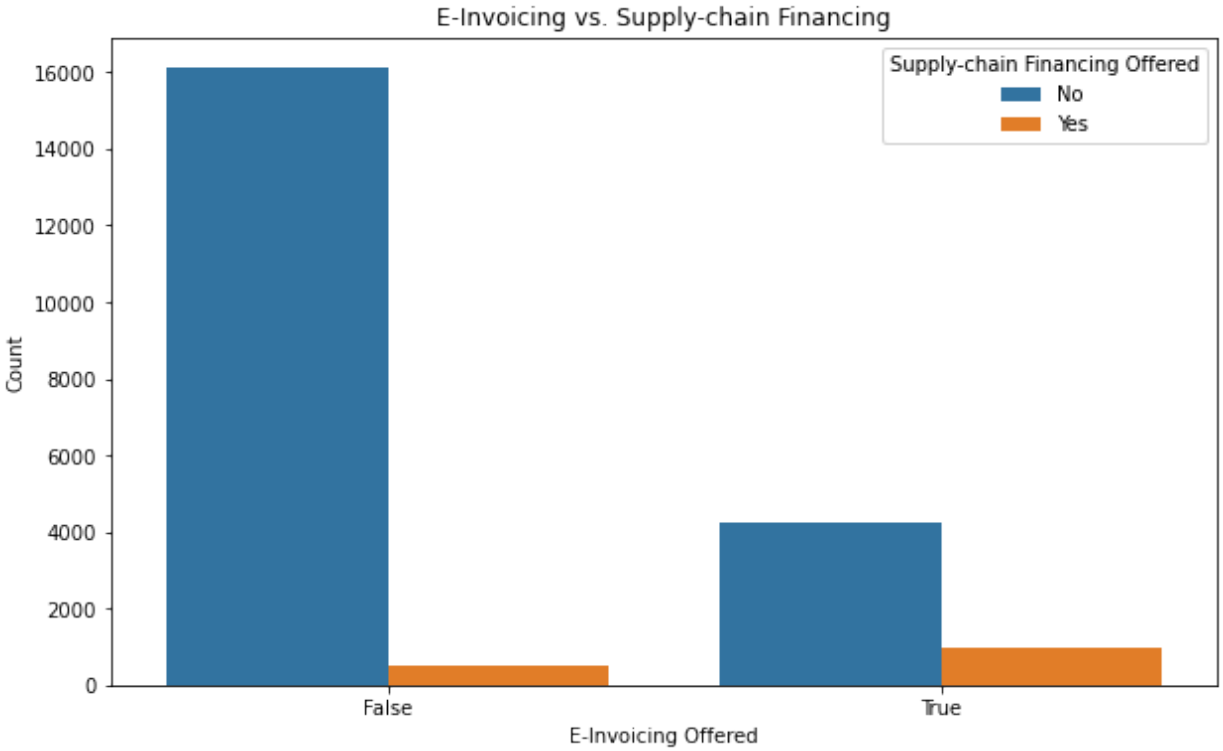
In [18]:
```
# 10. Correlation analysis for numeric columns
correlation_matrix = numeric_columns.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap', fontsize=16)
plt.show()
```



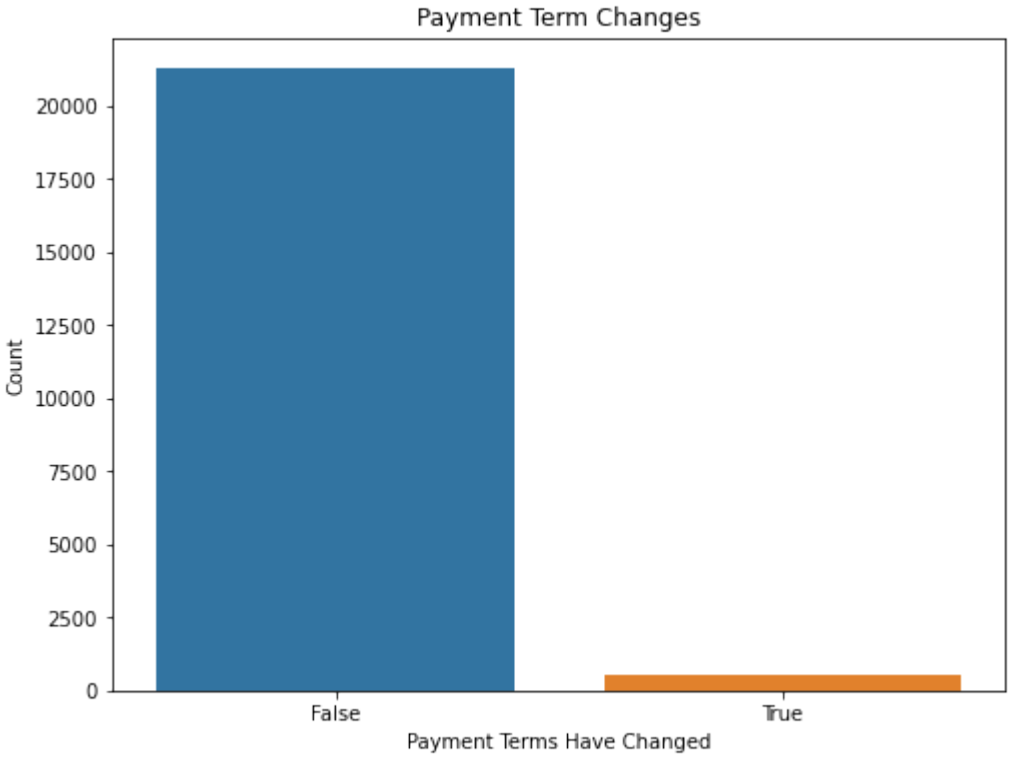
In [19]:
```
# Payment Code Participation
plt.figure(figsize=(8, 6))
sns.countplot(x='Participates in payment codes', data=data)
plt.title('Payment Code Participation')
plt.xlabel('Participates in Payment Codes')
plt.ylabel('Count')
plt.show()
```



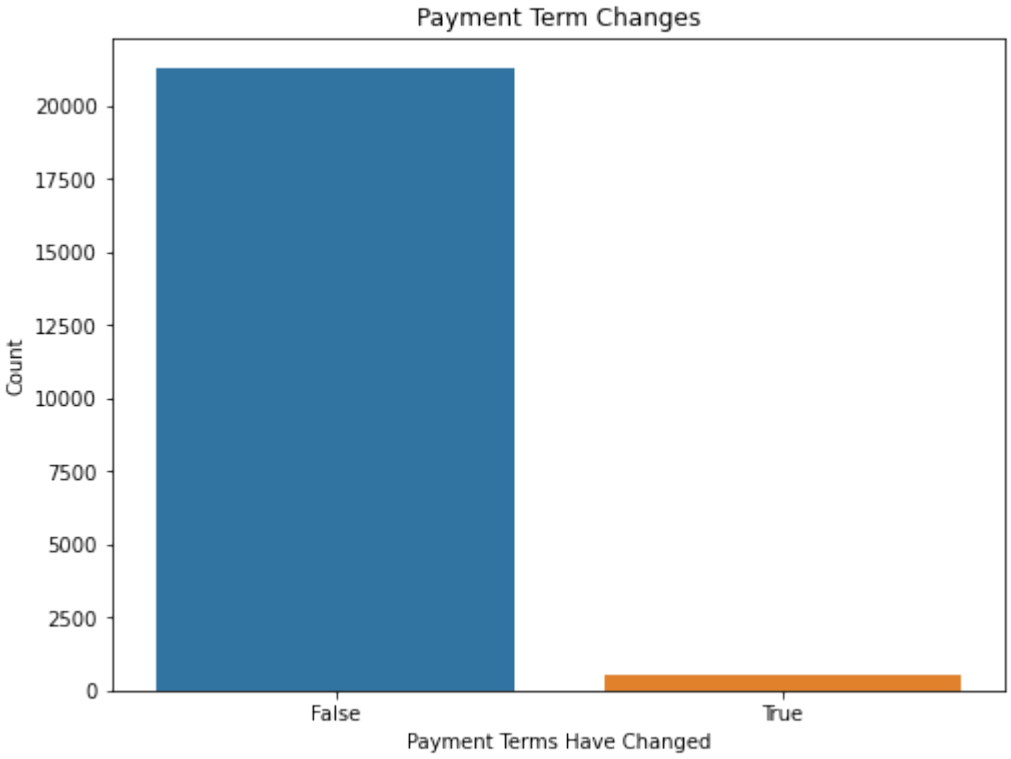
In [20]:
```
# E-Invoicing and Supply-chain Financing Comparison
plt.figure(figsize=(10, 6))
sns.countplot(x='E-Invoicing offered', hue='Supply-chain financing offered', data=data
plt.title('E-Invoicing vs. Supply-chain Financing')
plt.xlabel('E-Invoicing Offered')
plt.ylabel('Count')
plt.legend(title='Supply-chain Financing Offered', loc='upper right', labels=['No', 'Y
plt.show()
```



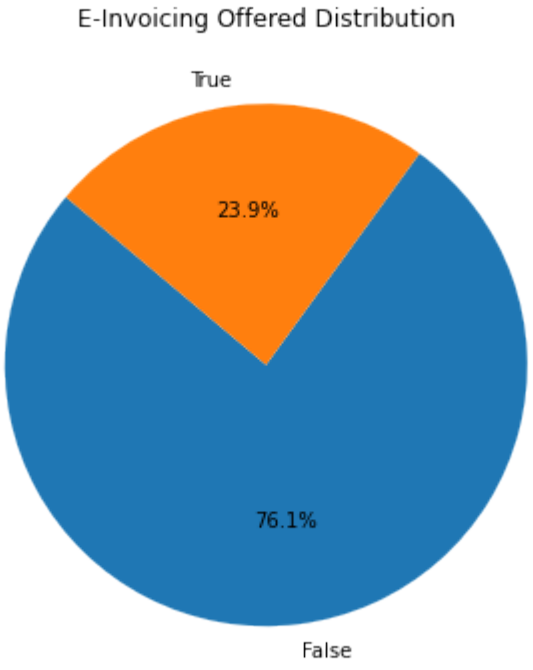
In [21]:
```
# Payment Term Changes
plt.figure(figsize=(8, 6))
sns.countplot(x='Payment terms have changed', data=data)
plt.title('Payment Term Changes')
plt.xlabel('Payment Terms Have Changed')
plt.ylabel('Count')
plt.show()
```



In [22]:
```
# Payment Term Changes
plt.figure(figsize=(8, 6))
sns.countplot(x='Payment terms have changed', data=data)
plt.title('Payment Term Changes')
plt.xlabel('Payment Terms Have Changed')
plt.ylabel('Count')
plt.show()
```



In [23]:
```
# Pie chart for the distribution of E-Invoicing offered
e_invoicing_distribution = data["E-Invoicing offered"].value_counts()
plt.figure(figsize=(6, 6))
plt.pie(e_invoicing_distribution, labels=e_invoicing_distribution.index, autopct='%1.1
plt.title('E-Invoicing Offered Distribution')
plt.show()
```



In [ ]: