

SampleData

January 25, 2025

```
[3]: import pandas as pd
import matplotlib.pyplot as plt

# Load the dataset
file_path = r'C:\Users\Corbi\generated_data.json'
data = pd.read_json(file_path)

# Create charts based on the dataset

# 1. Distribution of Age
plt.figure(figsize=(10, 6))
plt.hist(data['age'], bins=15, edgecolor='black')
plt.title('Distribution of Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.grid(axis='y')
plt.show()

# 2. Random Float by Currency
plt.figure(figsize=(12, 6))
currency_agg = data.groupby('currency')['random_float'].sum().
    ↪sort_values(ascending=False).head(10)
currency_agg.plot(kind='bar')
plt.title('Top 10 Currencies by Total Random Float')
plt.xlabel('Currency')
plt.ylabel('Total Random Float')
plt.xticks(rotation=45)
plt.show()

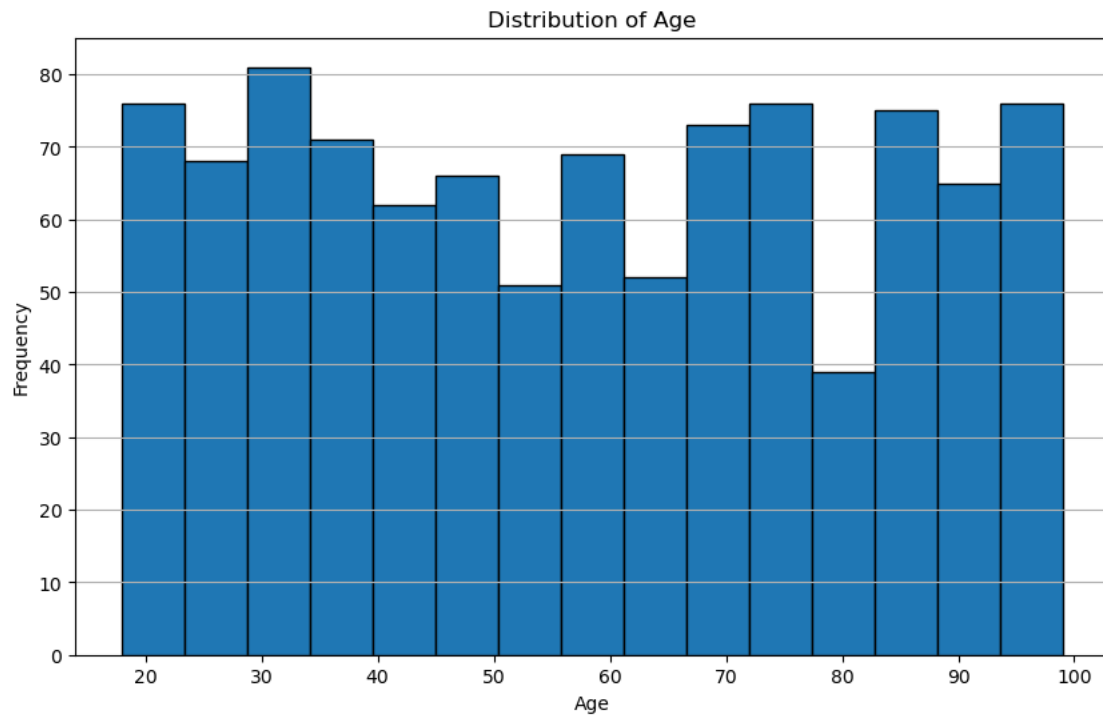
# 3. Percentage of True vs False in Boolean Column
boolean_counts = data['boolean'].value_counts()
plt.figure(figsize=(8, 6))
boolean_counts.plot(kind='pie', autopct='%1.1f%%', labels=['False', 'True'],
    ↪startangle=90, colors=['lightblue', 'orange'])
plt.title('Boolean Value Distribution')
plt.ylabel('')
plt.show()
```

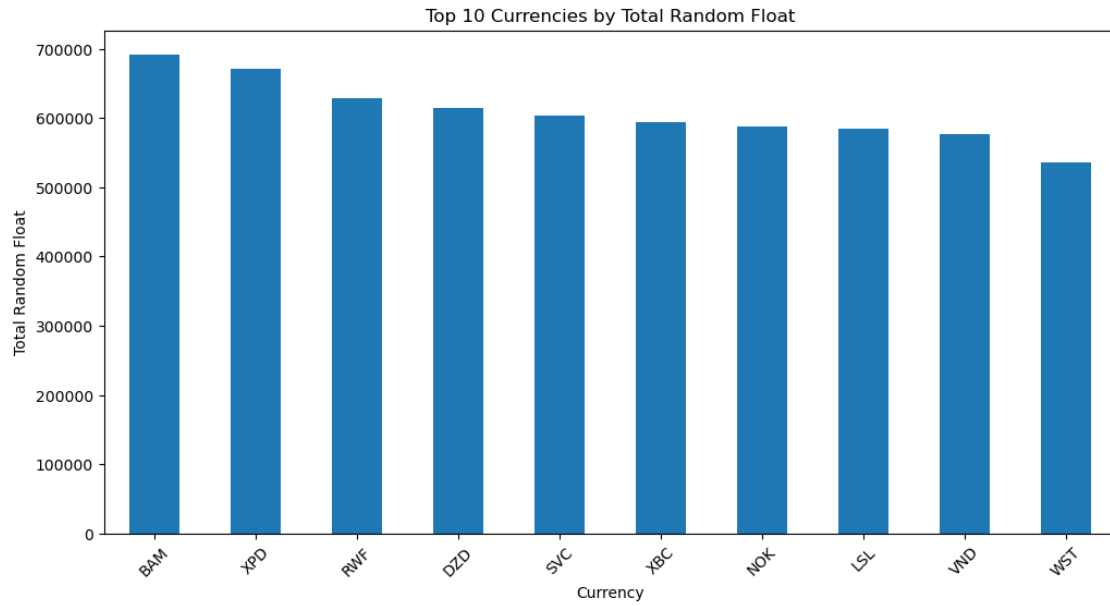
```

# 4. Top 10 Companies by Random Integer
plt.figure(figsize=(12, 6))
company_agg = data.groupby('company_name')['random_integer'].sum().
    ↪sort_values(ascending=False).head(10)
company_agg.plot(kind='bar', color='teal')
plt.title('Top 10 Companies by Total Random Integer')
plt.xlabel('Company Name')
plt.ylabel('Total Random Integer')
plt.xticks(rotation=45)
plt.show()

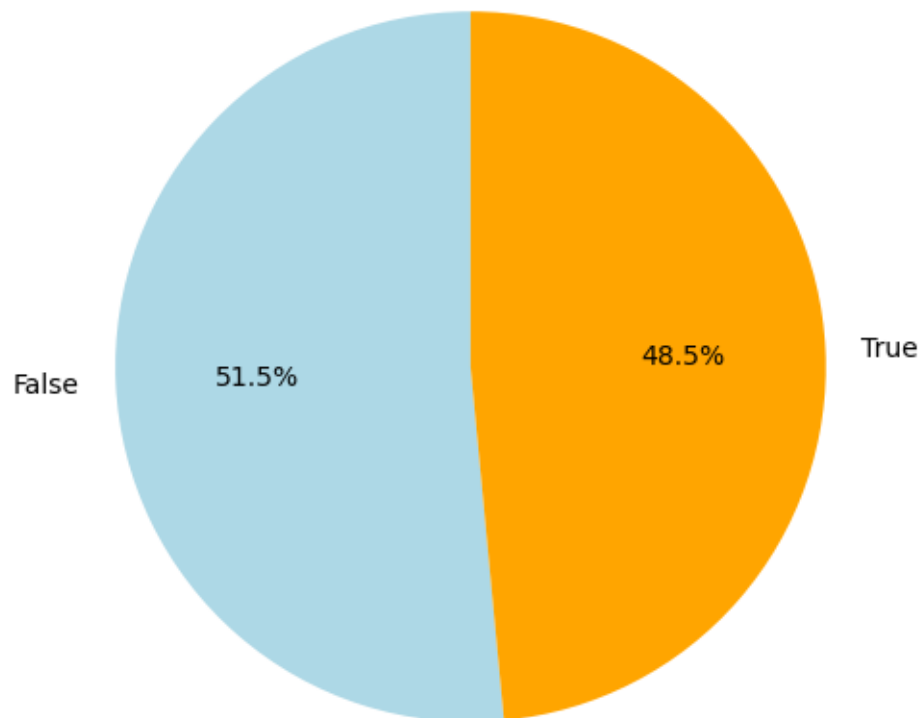
# 5. Latitude vs Longitude Scatter Plot
plt.figure(figsize=(10, 6))
plt.scatter(data['longitude'], data['latitude'], alpha=0.6, c='green')
plt.title('Scatter Plot of Latitude vs Longitude')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.grid(True)
plt.show()

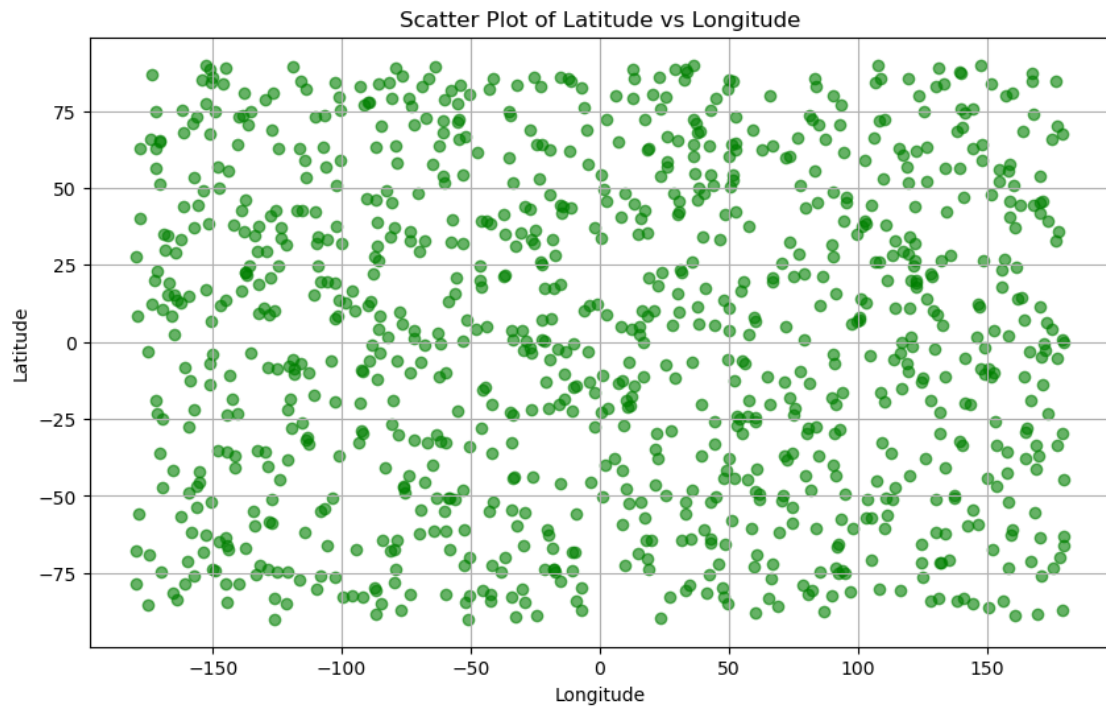
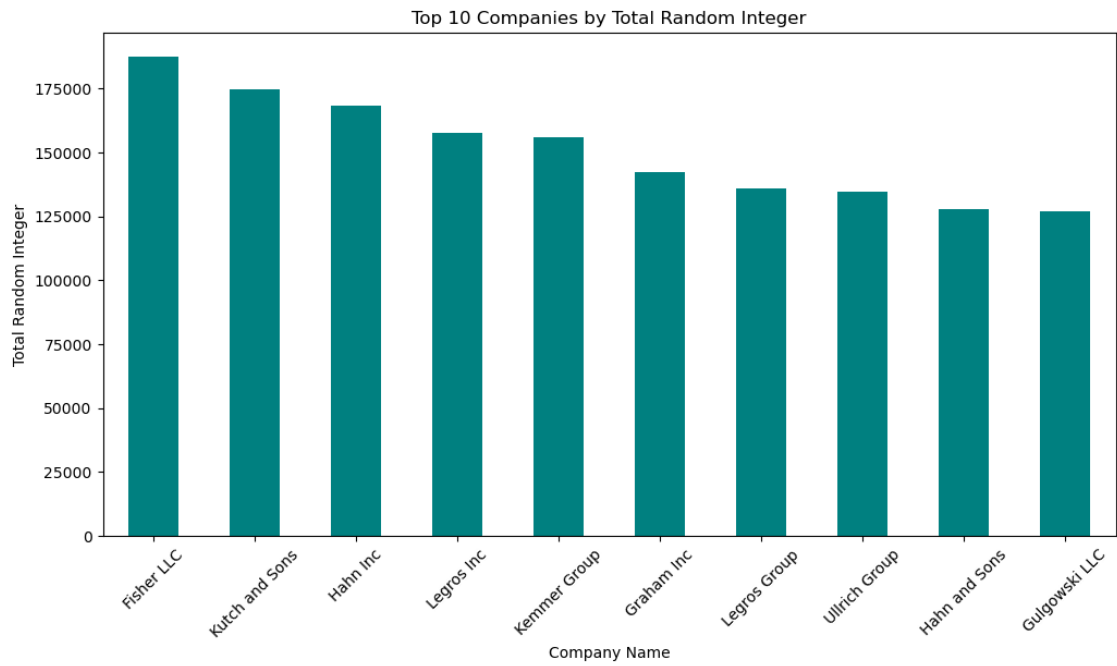
```





Boolean Value Distribution





[]: