In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [2]:
```
df = pd.read_csv("encounters.csv")
df.head()
```

Out[2]:

	Id	START	STOP	PATIENT	ORGANIZATION	PR
0	d0c40d10-8d87-447e-836e-99d26ad52ea5	2010-01-23T17:45:28Z	2010-01-23T18:10:28Z	034e9e3b-2def-4559-bb2a-7850888ae060	e002090d-4e92-300e-b41e-7d1f21dee4c6	e6 fd 0edb1
1	e88bc3a9-007c-405e-aabc-792a38f4aa2b	2012-01-23T17:45:28Z	2012-01-23T18:00:28Z	034e9e3b-2def-4559-bb2a-7850888ae060	772ee193-bb9f-30eb-9939-21e86c8e4da5	6f a5b 5bad2
2	8f104aa7-4ca9-4473-885a-bba2437df588	2001-05-01T15:02:18Z	2001-05-01T15:17:18Z	1d604da9-9a81-4ba9-80c2-de3375d59b40	5d4b9df1-93ae-3bc9-b680-03249990e558	af 31d 2867
3	b85c339a-6076-43ed-b9d0-9cf013dec49d	2011-07-28T15:02:18Z	2011-07-28T15:17:18Z	1d604da9-9a81-4ba9-80c2-de3375d59b40	3dc9bb2d-5d66-3e61-bf9a-e234c6433577	bb 262 d88e7
4	dae2b7cb-1316-4b78-954f-fa610a6c6d0e	2010-07-27T12:58:08Z	2010-07-27T13:28:08Z	10339b10-3cd1-4ac3-ac13-ec26728cb592	b03dba4f-892f-365c-bfd1-bfcfa7a98d5d	7e b84 15c42

In [3]:
```
df.info()
df = df.dropna(subset=["ENCOUNTERCLASS"])
```

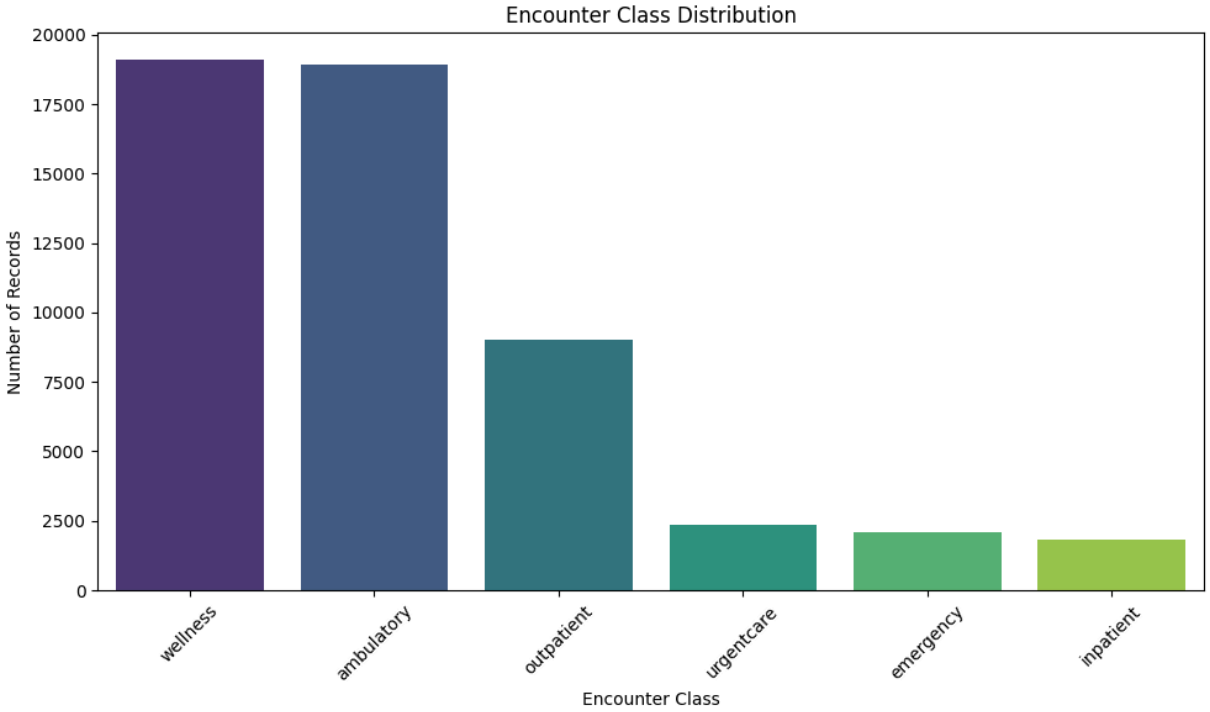
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 53346 entries, 0 to 53345
Data columns (total 15 columns):
 #   Column               Non-Null Count  Dtype  
---  ------               --------------  -----  
 0   Id                   53346 non-null  object 
 1   START                53346 non-null  object 
 2   STOP                 53346 non-null  object 
 3   PATIENT              53346 non-null  object 
 4   ORGANIZATION         53346 non-null  object 
 5   PROVIDER             53346 non-null  object 
 6   PAYER                53346 non-null  object 
 7   ENCOUNTERCLASS       53346 non-null  object 
 8   CODE                 53346 non-null  int64  
 9   DESCRIPTION          53346 non-null  object 
 10  BASE_ENCOUNTER_COST  53346 non-null  float64
 11  TOTAL_CLAIM_COST     53346 non-null  float64
 12  PAYER_COVERAGE       53346 non-null  float64
 13  REASONCODE           13777 non-null  float64
 14  REASONDESCRIPTION    13777 non-null  object 
dtypes: float64(4), int64(1), object(10)
memory usage: 6.1+ MB
```

In [4]:
```
encounter_counts = df['ENCOUNTERCLASS'].value_counts()
print(encounter_counts)
```

```
ENCOUNTERCLASS
wellness      19106
ambulatory    18936
outpatient     9003
urgentcare     2373
emergency      2090
inpatient      1838
Name: count, dtype: int64
```

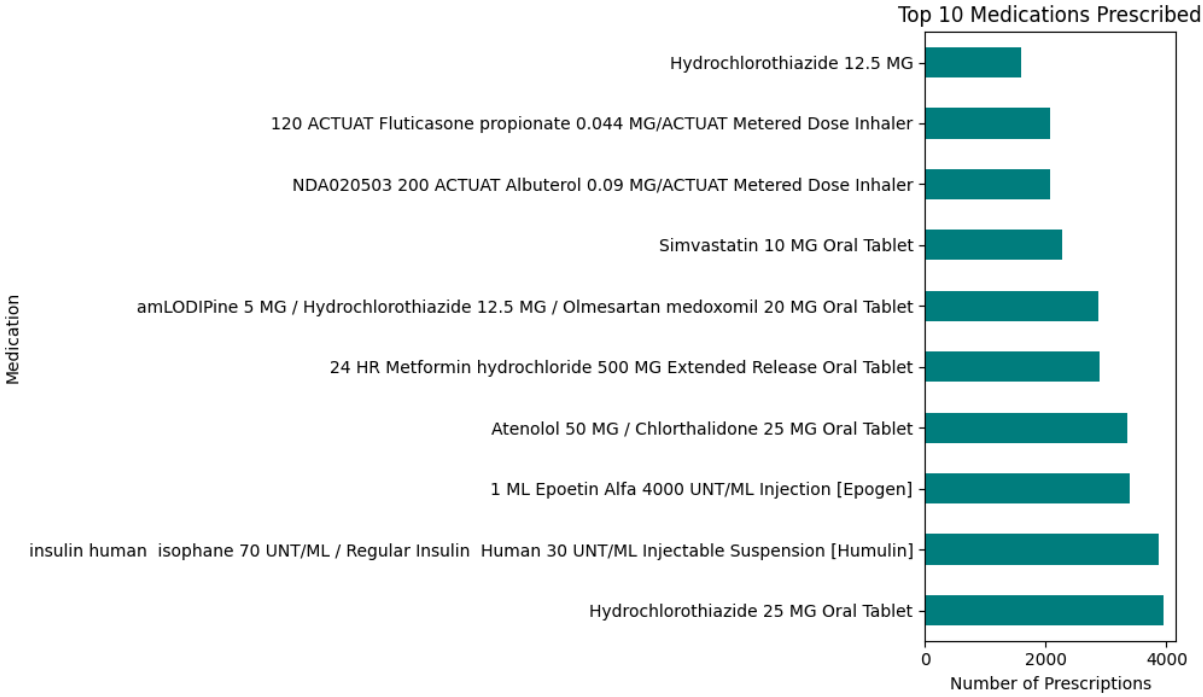
In [5]:
```
plt.figure(figsize=(10, 6))
sns.barplot(x=encounter_counts.index, y=encounter_counts.values, palette="vi
plt.title("Encounter Class Distribution")
plt.xlabel("Encounter Class")
plt.ylabel("Number of Records")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

```
/var/folders/lg/xp06b0710rl1ttt3_dpt43_80000gn/T/ipykernel_25064/2799454075.
py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed
in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the
same effect.

  sns.barplot(x=encounter_counts.index, y=encounter_counts.values, palette
="viridis")
```



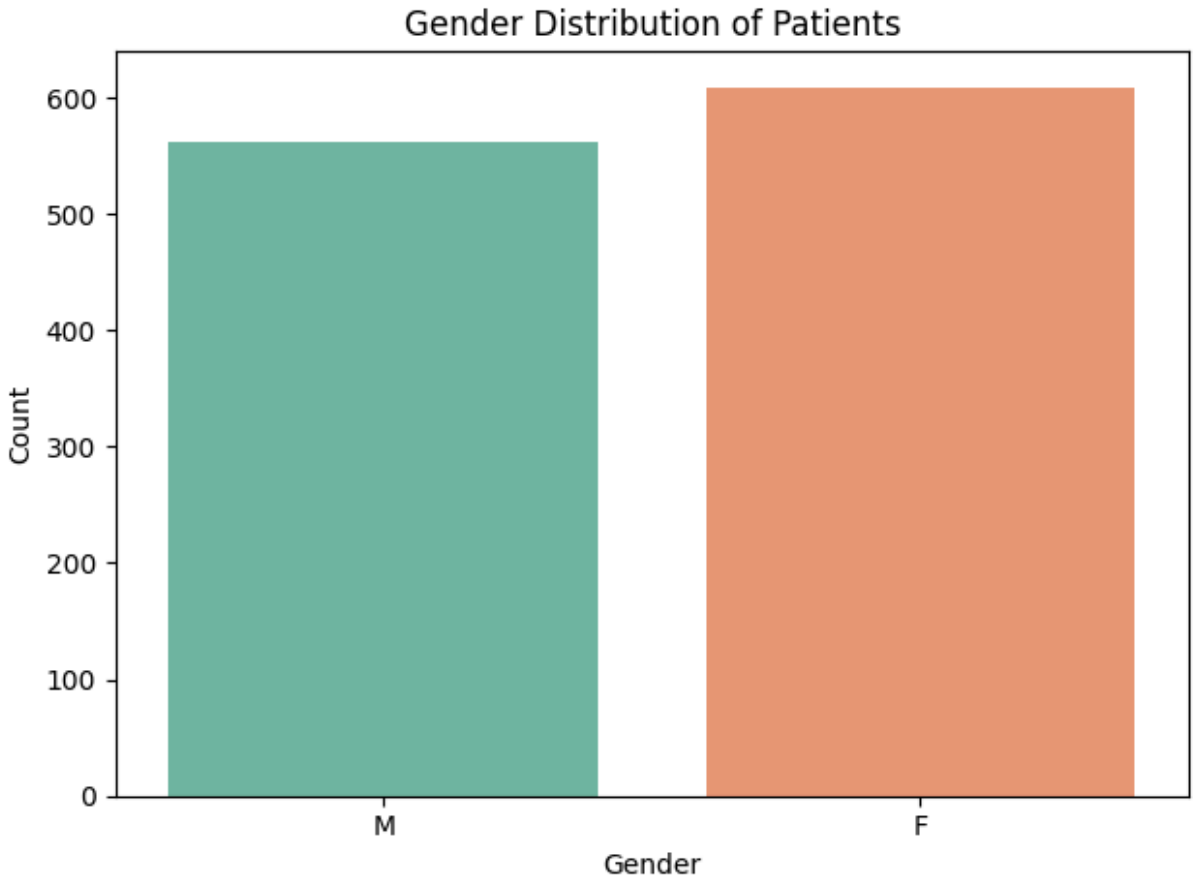
In [6]:
```
df.to_csv("cleaned_encounters.csv", index=False)
```

In [7]:
```
med_df = pd.read_csv("medications.csv")

# View top 5 rows (optional)
med_df.head()

# Plot
med_df['DESCRIPTION'].value_counts().head(10).plot(kind='barh', figsize=(10,
plt.title("Top 10 Medications Prescribed")
plt.xlabel("Number of Prescriptions")
plt.ylabel("Medication")
plt.tight_layout()
plt.show()
```

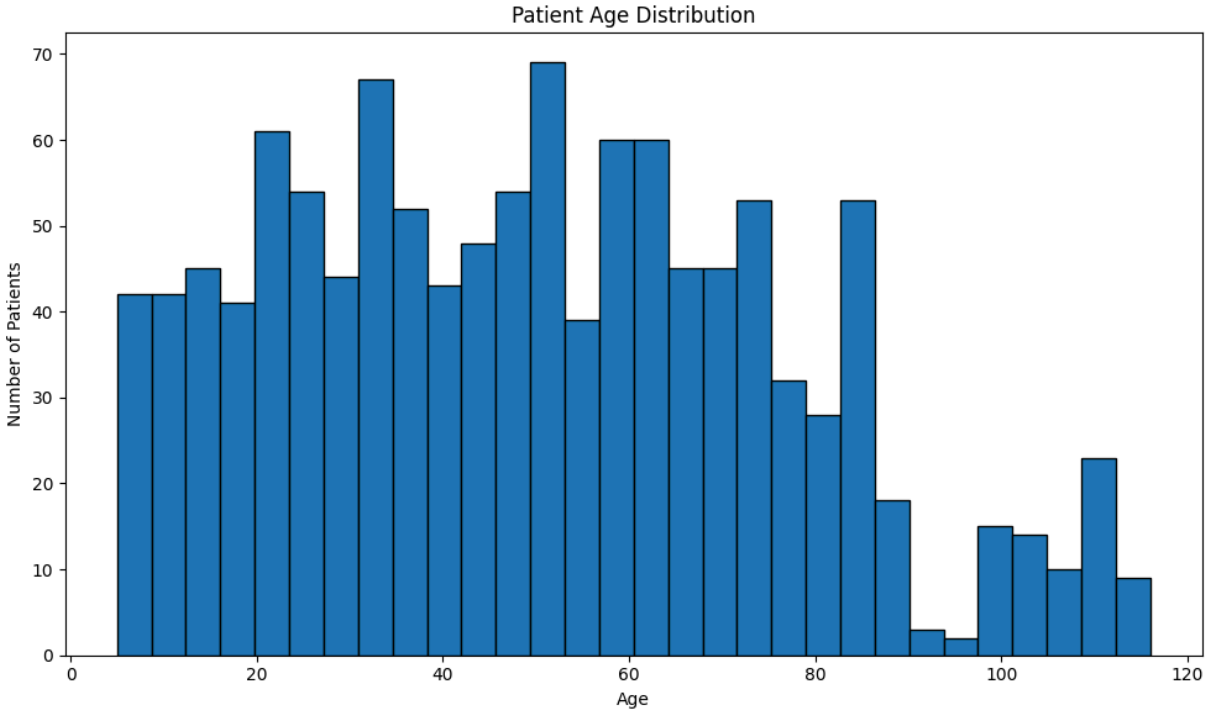


In [8]:
```
pat_df = pd.read_csv("patients.csv")

sns.countplot(data=pat_df, x="GENDER", palette="Set2")
plt.title("Gender Distribution of Patients")
plt.xlabel("Gender")
plt.ylabel("Count")
plt.tight_layout()
plt.show()
```

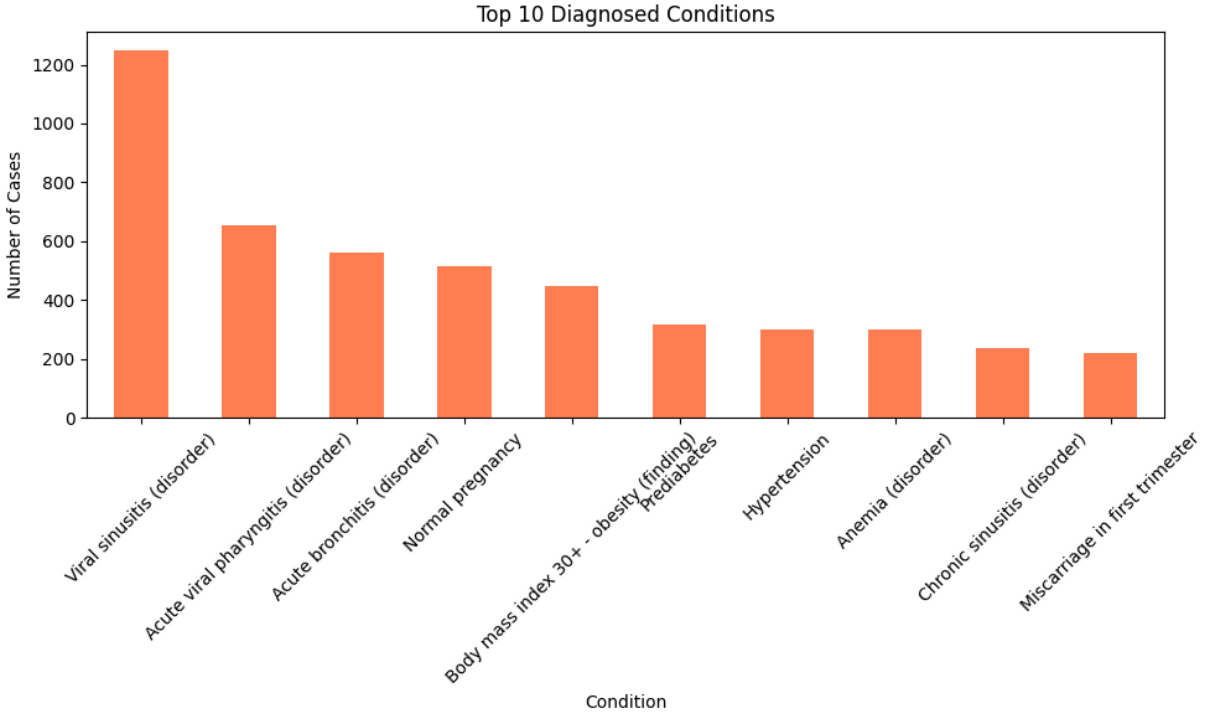
```
/var/folders/lg/xp06b0710rl1ttt3_dpt43_80000gn/T/ipykernel_25064/2309662404.
py:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed
in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the
same effect.

  sns.countplot(data=pat_df, x="GENDER", palette="Set2")
```



In [9]:
```
pat_df['BIRTHDATE'] = pd.to_datetime(pat_df['BIRTHDATE'])
pat_df['AGE'] = 2025 - pat_df['BIRTHDATE'].dt.year

pat_df['AGE'].plot(kind='hist', bins=30, edgecolor='black', figsize=(10,6))
plt.title("Patient Age Distribution")
plt.xlabel("Age")
plt.ylabel("Number of Patients")
plt.tight_layout()
plt.show()
```



In [10]:
```
cond_df = pd.read_csv("conditions.csv")

cond_df['DESCRIPTION'].value_counts().head(10).plot(kind='bar', figsize=(10,
plt.title("Top 10 Diagnosed Conditions")
plt.xlabel("Condition")
plt.ylabel("Number of Cases")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



In [ ]: